

## DOCUMENT RESUME

ED 389 722

TM 024 222

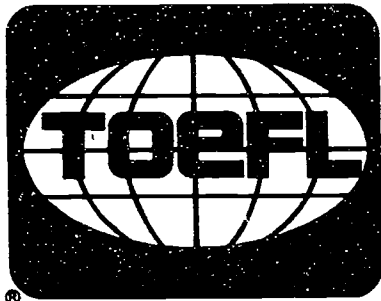
AUTHOR Hale, Gordon A.; And Others  
TITLE Multiple-Choice Cloze Items and the Test of English as a Foreign Language. TOEFL Research Reports 26.  
INSTITUTION Educational Testing Service, Princeton, N.J.  
REPORT NO ETS-RR-88-2  
PUB DATE Mar 88  
NOTE 149p.  
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC06 Plus Postage.  
DESCRIPTORS \*Adults; \*Cloze Procedure; \*English (Second Language); Estimation (Mathematics); Grammar; \*Item Response Theory; Language Tests; Listening Comprehension; \*Multiple Choice Tests; Reading Comprehension; Skills; Test Items; Vocabulary  
IDENTIFIERS Language Minorities; Nonnative Speakers; \*Test of English as a Foreign Language

## ABSTRACT

This study examined the relation of performance on the Test of English as a Foreign Language (TOEFL) to a widely used variant of the cloze procedure, the multiple choice (MC) cloze method. Examinees taking an operational TOEFL (n=11,290) were given three basic sections of the test along with a section containing prepared MC cloze items, and performance on each was examined for each of nine major language groups. Exploratory and confirmatory factor analyses suggested that, from a practical standpoint, TOEFL performance can be adequately described by two factors relating to listening comprehension and then to all other parts of the test. Examination of the MC cloze test showed that the total score was relatively reliable, and that it was possible to estimate item response theory parameters for the MC cloze items with reasonable accuracy. However, there was no strong empirical evidence that the items types within the MC cloze test reflected distinct skills. It appeared that skills associated with grammar, vocabulary, and reading comprehension are highly interrelated as assessed by the TOEFL and the MC cloze test. Five appendixes provide supplemental information about the analyses performed. (Contains 13 tables, 8 figures, and 37 references.) (SLD)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*



TEST OF ENGLISH AS A FOREIGN LANGUAGE

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

- ☒ This document has been reproduced as  
received from the person or organization  
originating it.
- ☐ Minor changes have been made to improve  
reproduction quality.
- Points of view or opinions stated in this docu-  
ment do not necessarily represent official  
OERI position or policy.

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

H. I. BRAUN

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC) "

# Research Reports

REPORT 26  
MARCH 1988

## Multiple-Choice Cloze Items and the Test of English as a Foreign Language

Gordon A. Hale  
Charles W. Stansfield  
Donald A. Rock  
Marilyn M. Hicks  
Frances A. Butler  
John W. Oller, Jr.

BEST COPY AVAILABLE



EDUCATIONAL TESTING SERVICE

The Test of English as a Foreign Language (TOEFL) was developed in 1963 by a National Council on the Testing of English as a Foreign Language, which was formed through the cooperative effort of over thirty organizations, public and private, that were concerned with testing the English proficiency of nonnative speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program and in 1973 a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the Graduate Record Examinations (GRE) Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; GRE Board members are associated with graduate education.

ETS administers the TOEFL program under the general direction of a Policy Council that was established by, and is affiliated with, the sponsoring organizations. Members of the Policy Council represent the College Board and the GRE Board and such institutions and agencies as graduate schools of business, junior and community colleges, nonprofit educational exchange agencies, and agencies of the United States government.

A continuing program of research related to TOEFL is carried out under the direction of the TOEFL Research Committee. Its six members include representatives of the Policy Council, the TOEFL Committee of Examiners, and distinguished English-as-a-second-language specialists from the academic community. Currently the committee meets twice yearly to review and approve proposals for test-related research and to set guidelines for the entire scope of the TOEFL research program. Members of the Research Committee serve three-year terms at the invitation of the Policy Council; the chair of the committee serves on the Policy Council.

Because the studies are specific to the test and the testing program, most of the actual research is conducted by ETS staff rather than by outside researchers. However, many projects require the cooperation of other institutions, particularly those with programs in the teaching of English as a foreign or second language. Representatives of such programs who are interested in participating in or conducting TOEFL-related research are invited to contact the TOEFL program office. Local research may sometimes require access to TOEFL data. In such cases, the program may provide this data following approval by the Research Committee. All TOEFL research projects must undergo appropriate ETS review to ascertain that the confidentiality of data will be protected.

Current (1987-88) members of the TOEFL Research Committee include the following:

Paul J. Angelis (chair)	Southern Illinois University
Patricia L. Carrell	Southern Illinois University
Lily Wong Fillmore	University of California at Berkeley
Russell G. Hamilton	Vanderbilt University
Frederick L. Jenks	Florida State University
Harold S. Madsen	Brigham Young University

Multiple-Choice Cloze Items and the  
Test of English as a Foreign Language

Gordon A. Hale, Educational Testing Service  
Charles W. Stansfield, Center for Applied Linguistics  
Donald A. Rock, Educational Testing Service  
Marilyn M. Hicks, Educational Testing Service  
Frances A. Butler, Consultant in language testing, Los Angeles  
John W. Oller, Jr., University of New Mexico

Educational Testing Service  
Princeton, New Jersey

RR 88-2

Educational Testing Service is an Equal Opportunity/Affirmative Action Employer.

Copyright © 1988 by Educational Testing Service. All rights reserved.

Unauthorized reproduction in whole or in part is prohibited.

TOEFL, the TOEFL logo, ETS, and the ETS logo are trademarks of Educational Testing Service, registered in the USA and in many other countries.

### Acknowledgments

The authors wish to express their gratitude to the following ETS staff members, who contributed generously of their time to this project:

Tom Jirele, for the principal programming work

David Sands and Jennifer Nelson, for additional programming

Marian Helms, for secretarial assistance

Barbara Suomi, for assistance in final preparation of the cloze items

Robert Stellman, Willem Spits, and Barbara Mathews, for help in coordinating the test preparation

Lawrence Stricker, Walter Emmerich, Jackie Ross, and Denise Asfar, for a critical reading of an earlier draft of this manuscript.

For additional suggestions regarding the manuscript, the authors express their thanks to Lih Shing Wang of the Center for Applied Linguistics.

The authors wish to extend their sincere appreciation to Lyle Bachman of the University of Illinois and to Marianne Celce-Murcia of the University of California at Los Angeles for their participation in classifying the cloze items.

The authors gratefully acknowledge the contributions made by the staff members of the institutions at which the multiple-choice cloze items were pretested: Louis Levi and Richard Tudisco of Columbia University; Maria Jacobson, Beverly Spencer, and Lena McTigue of Hunter College, City University of New York; and Karen Medina and Ravee Raghavan of Rutgers University.

Finally, the authors are indebted to the TOEFL Research Committee for its support of the project.

## Abstract

This study examined the relation of TOEFL<sup>®</sup> performance to a widely used variant of the cloze procedure--the multiple-choice (MC) cloze method. A main objective was to determine if categories of MC cloze items could be identified that related differentially to the various parts of the TOEFL. MC cloze items were prepared and classified according to whether the involvement of reading comprehension, as defined by sensitivity to long-range textual constraints, was primary or secondary. For two categories, reading comprehension was primary and knowledge of grammar or vocabulary was secondary, and for two other categories knowledge of grammar or vocabulary was primary and reading comprehension secondary.

Examinees taking an operational TOEFL at domestic test centers were given the three basic sections of the test along with a fourth section containing the MC cloze items. Performance was examined for each of nine major language groups.

Exploratory and confirmatory factor analyses for the basic TOEFL were performed first, to provide a basis for relating the MC cloze items to the TOEFL structure. These factor analyses suggested that, from a practical standpoint, TOEFL performance can be adequately described by just two factors, which relate to (a) Listening Comprehension, and (b) all other parts of the test--Structure, Written Expression, Vocabulary, and Reading Comprehension.

Examination of the MC cloze test showed that the total MC cloze score was relatively reliable and that it was possible to estimate item response theory parameters for the MC cloze items with reasonable accuracy. Thus, the development of the MC cloze items was successful in these respects. However, the correlations among scores for the four MC cloze item categories were approximately as high as their reliabilities, thus providing no strong empirical evidence that the item types within the MC cloze test reflected distinct skills.

Correlational analyses related the four MC cloze categories to the five parts of the TOEFL. These analyses revealed a slight tendency for MC cloze items that involved a combination of grammar and reading to relate more highly to the Structure and Written Expression parts of the TOEFL than the other parts, and for MC cloze items that involved a combination of vocabulary and reading to relate more highly to the Vocabulary and Reading Comprehension parts of the TOEFL than

the other parts. Although this pattern was relatively consistent across language groups, however, the differences among correlations were not substantial enough to be of practical importance.

Multiple regression analyses were performed, using total MC cloze score as the dependent variable and the five TOEFL parts as independent variables. The resulting multiple Rs were mostly in the lower to upper .90s, suggesting that total MC cloze performance can be predicted from TOEFL performance with a relatively high degree of accuracy.

In general, the study provided no evidence that distinct skills are measured by the nonlistening parts of the TOEFL or by the four categories of MC cloze items. It would appear that the skills associated with grammar, vocabulary, and reading comprehension are highly interrelated, as assessed by the TOEFL and the MC cloze test.



## Table of Contents

	<u>Page</u>
Introduction.....	1
Overview of Selected Literature.....	2
Objectives of the Present Research.....	7
Method.....	9
Subjects.....	9
Materials.....	9
Procedure.....	12
Results.....	13
Method of Analysis.....	13
Analyses of the TOEFL.....	13
Analyses of MC Cloze Items.....	27
The Relation of MC Cloze Items to the TOEFL.....	29
IRT Parameter Estimation, Test of Unidimensionality, and Equating of MC Cloze Items.....	36
Discussion.....	51
Practical Implications.....	53
References.....	55
Appendices.....	59
A. TOEFL Test Form Used in the Study, Including Fourth, Cloze Section.....	59
B. Categories of the MC Cloze Items.....	77
C. Flyer Sent to Examinees in Advance of Test Administration.....	81
D. Varimax Factors Obtained in Two-, Three-, and Five-Factor Solutions for the TOEFL.....	87
E. Confirmatory Factor Analyses: Factor Loadings for One-, Two-, Three-, and Five-Factor Solutions.....	107

# List of Tables

<u>Table No.</u>		<u>Page</u>
1	Means and Standard Deviations of Raw Scores on the TOEFL for Each Language Group.....	15
2	Correlations Among TOEFL Subscores and Reliabilities..	16
3	First Six Eigenvalues from Principal Components Factor Analysis of the TOEFL and Cumulative Percentage of Total Variance Accounted for, by Language Group....	19
4	Indices from Confirmatory Factor Analysis of the TOEFL for One-, Two-, Three-, and Five-Factor Solutions.....	22
5	Correlations among Factors in Two- and Three-Factor Confirmatory Factor Analyses of the TOEFL.....	24
6	Correlations among Factors in Five-Factor Confirmatory Factor Analysis of the TOEFL.....	25
7	Means and Standard Deviations of Cloze Scores.....	28
8	Correlations among Cloze Scores and Reliabilities.....	30
9	Correlations of Cloze and TOEFL Scores.....	31
10	Correlations of Cloze and TOEFL Scores Corrected for Unreliability.....	33
11	Average Chi-Square Values for Cloze and Operational Items.....	42
12	Frequency Distributions of a parameters.....	46
13	Frequency Distributions of b parameters.....	47

## List of Figures

<u>Figure No.</u>		<u>Page</u>
1	Examples of item ability regressions for cloze items scaled with Section 2.....	39
2	Examples of item ability regressions for cloze items scaled with Section 3.....	40
3	Item ability regressions for cloze item #4 scaled with Sections 2.....	41
4	Item ability regression for cloze item #4 scaled with Section 3.....	41
5	Principal axis plot of CL0 and CL2.....	44
6	Principal axis plot of CL0 and CL3.....	45
7	Relative efficiency for the test composed of cloze items.....	49
8	Typical relative efficiency curve for TOEFL Section 3.....	50

## Introduction

Recent years have seen an increased interest in the cloze method of testing. Originally developed as a means of determining the readability of text (Taylor, 1953), the cloze procedure has since been considered by many to be an effective means of assessing second language proficiency (e.g., Anderson, 1976; Oller, 1979). In the cloze procedure, examinees encounter a segment of text from which words have been deleted and replaced by blanks, and they must indicate the word (or other segment) that best fills in the blank space.

Part of the appeal of cloze testing lies in the fact that it is regarded by many as an "integrative" method of assessment. In this respect, it contrasts with "discrete-point" methods, which are supposed to deal with a single component of a language and a single skill at a time. The cloze procedure deals with several linguistic components at once, focusing more on language use, and typically requiring the examinee to read and comprehend a substantial amount of discourse (e.g., Carroll, 1961; Oller, 1979).<sup>1</sup>

The Test of English as a Foreign Language (TOEFL) is a widely used measure whose primary purpose is to assess the English proficiency of foreign applicants to universities in the United States and Canada. Despite the test's validity, generally shown through correlations with other batteries of English proficiency tests (Educational Testing Service, 1987), research is continually being done to further refine the test. Among the issues for research is whether there may be effective ways of assessing reading comprehension in addition to the traditional method used in the TOEFL (i.e., text passages followed by questions about them). Among reasons for this, a previous study (Swinton & Powers, 1980) implied that the Vocabulary and Reading Comprehension section may not tap a unitary dimension for certain language groups; yet the function of this test section is to measure a single dimension related to reading ability (where both vocabulary and reading comprehension items are intended to tap this ability). Therefore, it was of interest to consider the possibility of adding one or more new reading-related item types, perhaps in place of vocabulary items, to ensure more homogeneous measurement of reading ability across language groups.

It was believed that the cloze procedure might prove useful in this regard. Research reviewed below suggests its potential value in assessing processes related to reading comprehension. And, although the cloze method is not limited to assessment of reading comprehension, it was assumed feasible to identify cloze items that are

---

<sup>1</sup>In this discussion, "cloze testing" refers to the standard random deletion cloze method, where typically every *n*th word is deleted. The answer format may be open ended or multiple choice. There are, however, variations on the standard procedure that can be more discrete point in orientation, such as cloze passages in which only articles or only prepositions are deleted.

specifically targeted at assessment of reading comprehension.

More specifically, the present study examined a multiple-choice version of the cloze method. In the traditional cloze procedure, the examinee is required to fill in the blanks by writing in the words that best fit the text. Where mass administration and scoring are required, however, as with the TOEFL, this procedure would be impractical. A feasible alternative is the multiple-choice (MC) cloze format, in which examinees choose, from among several alternatives, the response that best fills in the blank for each deleted word (cf. Bensoussan, 1983; Guthrie, 1973; Hinofotis & Snow, 1978; Jonz, 1976.) The multiple-choice cloze format, although lacking the free-response feature of the standard cloze procedure, has many aspects of the cloze method that make it desirable for consideration, as suggested in the following overview of the literature.

### Overview of Selected Literature

This section summarizes findings in the literature on cloze testing that are directly related to the present study. The emphasis is on studies that have employed a multiple-choice version of the cloze test. However, to provide background information, the standard cloze procedure is briefly considered first.

Standard cloze procedure. In the traditional form of the standard (or "completion") cloze test, every *n*th (e.g., every fifth or seventh) word is deleted (cf. Taylor, 1953). In a relatively common variant, deletion occurs at irregular intervals, usually to allow for deletion of specific types of words. The correct response can be the exact deleted word, or it can be a reasonable substitute, with either alternative producing acceptable results (e.g., Hinofotis, 1978; Irvine, Atai, & Oller, 1974).

Research using the standard cloze procedure suggests that it is an effective measure of English proficiency, both with native English speakers (e.g., Anderson, 1976; Bormuth, 1966; O'Reilly & Streeter, 1977) and with nonnative English speakers (e.g., Anderson, 1976; Friedman, 1964; Oller, 1979). Several of these authors have argued that the cloze procedure provides information that relates particularly to reading comprehension, although others (e.g., Alderson, 1983) have argued that the procedure measures other aspects of proficiency such as grammar and vocabulary, to a greater degree. Whatever the prevailing view, however, it seems reasonable to conclude that reading ability is among the key components of English proficiency tapped by the cloze procedure. A plausible hypothesis, then, is that one can look at each individual item in a cloze test and identify items that are especially likely to tap processes related to reading comprehension, as defined by sensitivity to long-range textual constraints (cf. Bachman, 1985).

Multiple-choice cloze test--native English speakers. The multiple-choice version of the cloze test has been less thoroughly researched than the completion cloze test, but the research that has been done is sufficient to allow certain hypotheses to be drawn. One group of studies has involved native English speakers--usually elementary school children in the United States.

O'Reilly and his associates have conducted research with native English-speaking elementary school children (e.g., O'Reilly & Schuder, 1977; O'Reilly & Streeter, 1977). O'Reilly and Streeter performed a factor analysis of scores on various tests, including the MC cloze test, the California Achievement Test (CAT), a test of intelligence, and other measures. Two major factors were identified in this analysis. One factor apparently involved "literal comprehension," or "the apprehension of strictly literal meanings contained in sentences and phrases as measured by reading tests that focus on factual questions, questions about explicit details, and questions about interpretation of meanings within the context of isolated sentences and phrases" (O'Reilly & Streeter, 1977, p. 67). Relatively high loadings on this factor were found for the MC cloze test as well as for standard reading subtests tapping the above-mentioned types of processes. A second factor, on which the MC cloze test showed only a minor loading, was one related to inferential reasoning and intelligence; the highest loadings on this factor were found for the intelligence test and for the inferential subtests of the CAT. O'Reilly and Streeter concluded that the MC cloze format is "in part a measure of a restricted form of reading comprehension that is essentially independent of IQ" (p. 67).

In other research involving MC cloze tests with native English speakers, Guthrie (1973) found correlations in the low to mid .80s with scores on the Gates-MacGinitie Reading Test for normal and reading disabled readers of two age groups; and Cranney (1973) found a correlation of .52 with scores on the Cooperative Reading Test for college sophomores. (Cranney also found a correlation of .52 between completion cloze and Cooperative Reading Test scores.) These results provide at least modest additional support for the view that performance on an MC cloze test relates to reading comprehension in native English speakers.

Multiple choice cloze test--nonnative English speakers. In the several studies of MC cloze tests that have been conducted with nonnative English speakers, one issue of concern has been whether the score on an MC cloze test relates to the score on a completion cloze test. Hinofotis and Snow (1978) studied 66 students entering a university ESL program. They obtained a correlation of .54 between MC cloze and completion cloze tests, using exact word scoring of the latter test, and a correlation of .59 using acceptable word scoring. Pike (1979) studied TOEFL examinees from Peru, Chile, and Japan (Ns per country ranging from 98 to 199), who were administered several measures within a few days after taking the TOEFL, including an MC cloze and a completion cloze test. Correlations between MC cloze and

completion cloze tests ranged from .70 (Japan) to .89 (Peru); correlations corrected for unreliability ranged from .86 to .99. Variation among countries was attributed partly to differences in the range of proficiency, with the greatest correlations between tests observed for the Peruvian students, for whom the greatest range in test scores was observed. An implication of these findings, then, is that an MC cloze test and a completion cloze test may be measuring similar processes.

Concurrent validity data from these and other studies also bear on the question of similarity in processes measured by MC and completion cloze tests. Hinofotis and Snow (1978) correlated each type of test with the placement test of the Center for English as a Second Language (CESL) at Southern Illinois University. For all students combined (i.e., those who had the MC cloze test first, and those who had the completion cloze test first), correlations of the MC cloze and completion cloze scores with total CESL test scores were .63 and .71, respectively. The difference between these correlations was not significant, leading the authors tentatively to hypothesize that the MC and completion cloze tests provide similar information. (The authors also note, however, that "the depressed variability in the sample prevents us from concluding that an MC cloze test cannot be substituted for a more complicated testing procedure" [p. 133].)

Brown (1980) examined the relation of both MC and completion cloze tests to the UCLA ESL placement examination (ESLPE) and observed a correlation of .89 for the MC cloze test ( $N = 57$ ) and a similar correlation for the completion cloze test ( $N = 55$ ) (.88 for exact word scoring, and .90 for acceptable word scoring.) Scholz and Scholz (1981) examined the concurrent validity of MC and completion cloze tests for Chinese students of the English language, measured against a combination of five English proficiency subtests: the Structure and Listening subtests of the Comprehensive English Language Test (CELT), and the Reading Comprehension, Vocabulary, and Writing Ability subtests of the five-part TOEFL. They concluded that the concurrent validity was about the same for the completion cloze and two versions of the MC cloze test, one of which involved rational choice (i.e., selection by teachers) of distractors, and the other of which involved empirical selection of distractors, using errors made by a previous sample of students. (The latter method was also used by Hinofotis and Snow and by Brown; the method of selecting distractors will be discussed further below.) A reasonable conclusion from the above studies, then, is that the MC cloze test and completion cloze test measure similar processes, as evidenced by the fact that they correlate at about the same level with criterion measures of English language proficiency.

Correlations of MC cloze test with criterion subtests. It is useful to examine the data from these studies further, in an effort to determine the kinds of processes that are measured by MC cloze tests. In particular, it is of value to focus on studies involving nonnative English speakers and to determine the degree to which MC cloze tests



correlate with each of several subtests of English proficiency tests. Hinofotis and Snow (1978), examining subscores of the CESL placement test, found that the correlation with an MC cloze test was .52 for the CESL listening subscore, .53 for the reading subscore, and .62 for the English structure subscore. Jonz (1976) found that an MC cloze test showed the following correlations with subscores of the Virginia Placement Test: Reading, .61; Composition, .80; Aural, .29; Vocabulary, .54; and Structure, .70. Baldauf and Propst (1979) administered MC cloze tests and the Gates-MacGinitie (GM) reading test to fourth- and fifth-grade ESL students, and found correlations ranging from .68 to .83 between the MC cloze method and the GM Vocabulary subscore (the magnitude of the correlation depending on the students' grade and the method by which distractors were selected for the MC cloze test), and correlations ranging from .69 to .76 for the GM Comprehension subscore. In these studies, then, the MC cloze score appeared to correlate moderately with scores in reading comprehension, English structure (where examined), and vocabulary (where examined).

A few studies have related MC cloze performance to subscores of the TOEFL. Manning (1987), in a study that focused primarily on another form of cloze procedure (the cloze-elide method), examined the relation of MC cloze items to the TOEFL. He found that MC cloze scores correlated .65, .73, and .73, respectively, with (a) Listening Comprehension, (b) Structure and Written Expression, and (c) Vocabulary and Reading Comprehension. In two other studies, an MC cloze test was administered along with subtests of the five-part TOEFL. Pike (1979) observed correlations with MC cloze scores, corrected for unreliability, that ranged from .63 to .88 for Listening Comprehension; .71 to .87 for English Structure; .80 to .98 for Vocabulary; .93 to .95 for Reading Comprehension; and .76 to .90 for Writing Ability, the magnitude of the correlation depending on the examinees' native country. (Medians across countries = .69, .83, .86, .94, and .88, respectively, for the five subtests.) Scholz and Scholz (1981) found that, for the MC cloze test involving teacher-selected distractors (which yielded the highest validity coefficients among various methods of selecting distractors), correlations of the MC cloze test with each of the three TOEFL subtests studied were: Reading Comprehension for two different types of prose passages, .55 and .69; Vocabulary, .53 and .66; and Writing Ability, .54 and .59. Thus, in these two studies involving the TOEFL, as well as in the above studies involving other criterion measures, correlations with the MC cloze scores were moderately high for the Reading Comprehension scores, as well as for the Vocabulary and Writing Ability scores.

A general conclusion that may be derived from these studies is that the MC cloze procedure taps processes relating to several aspects of English proficiency, including reading comprehension, vocabulary, and receptive aspects of writing ability.

Empirical vs. rational methods of creating distractors. Different methods of creating items in the MC cloze test--particularly, different methods of determining the distractors--could be a factor



underlying whatever differences in results may have been observed across studies. Two principal methods have been used in research conducted to date. One--the empirical method--is to use common errors observed via pretesting with a completion cloze test; this is the method used in the studies cited above by Brown, Cranney, Hinofotis and Snow, Jonz, and Scholz and Scholz. The other--the rational selection method--is to have specialists select distractors based on examination of the passage and on consideration of possible sources of examinee error. This method has been employed in the above-mentioned studies by Baldauf and Propst, Guthrie, and Scholz and Scholz, and also in a study by Bensoussan and Ramraz (1984). (Another method, employed by O'Reilly and associates, is computerized random identification of distractors from a list of common words.)

One of the studies discussed above (Scholz and Scholz, 1981) employed both empirical and rational methods of distractor selection and thus obtained data that would allow comparison of the two. Examining the concurrent validity of the MC cloze test against a combination of English proficiency subtests as a criterion (see above), they found that the rational method of distractor selection for the MC cloze test yielded correlations of .62 and .74 (for two passages) with the criterion. (Note that these correlations were comparable to those obtained with completion cloze items-- .67 and .64.) In contrast, the correlations were lower for distractors derived empirically through (a) use of common errors made by pretest examinees with another native language (.49 and .66) and (b) use of common errors by examinees with the same native language (.45 and .58). Furthermore, correlations with the TOEFL subtests were higher in every case for MC cloze items based on rational selection of distractors than for items based on empirical selection. This is the only study located that compared empirical and rational methods of distractor selection; but to the extent it is possible to generalize from these data, it appears that rational selection of distractors can be at least as effective as empirical selection.

Summary of evidence on MC cloze test. To summarize, it appears that an MC cloze test taps processes related to several aspects of English proficiency, including reading comprehension, vocabulary, grammar, and receptive aspects of writing ability. Although most studies of the MC cloze test have used distractors selected empirically, based on errors made on a completion cloze test, the one study identified that compared various methods showed that rational selection of distractors by specialists can be as effective as empirical selection.

An interpretation. That MC cloze items apparently tap processes related to several aspects of English proficiency may be due in part to the fact that the typical MC cloze test contains a number of different types of items. For example, some items can be regarded as tapping knowledge of vocabulary, without need for comprehension across clausal boundaries. An example of such an item, taken from the Secondary Level English Proficiency Test, is the following: "The

early humans discovered the \_\_\_\_\_ of living and working together in groups. (Response alternatives: (a) behaviors, (b) advantages, (c) theories, (d) arrangements)." Other item categories can tap examinees' understanding of structural or semantic aspects of text. Furthermore, among items that tap semantic aspects, some items may require understanding of only a small portion of text immediately surrounding the deletion, while other items may require understanding of a much larger portion of text.

It is hypothesized, then, that one can identify different categories of MC cloze items that tap different components of English proficiency, and that those items that primarily involve reading comprehension, as defined by sensitivity to long-range textual constraints, can be specifically targeted for special attention. This assumption is consistent with the arguments put forward by Bachman (1985) in a study using the standard cloze procedure and by Bensoussan and Ramraz (1984) in a study using the MC cloze method.

### Objectives of the Present Research

The purpose of the present research was to determine how various categories of MC cloze items relate to the structure of the TOEFL. More specifically, the objectives were (a) to develop an MC cloze test in which the items could be classified according to the areas measured on the nonaural parts of the TOEFL--grammar, vocabulary, and reading comprehension--and (b) to determine the relation of each category of MC cloze item to each part of the TOEFL. It was assumed that the MC cloze category that was thought to tap a particular aspect of language proficiency would relate most highly to the TOEFL item type designed to measure the same aspect. Thus, MC cloze items classified as tapping grammar might be expected to relate most strongly to TOEFL items in the Structure and Written Expression section; MC cloze items classified as tapping vocabulary might relate most strongly to TOEFL Vocabulary items; and MC cloze items regarded as tapping reading comprehension might relate most strongly to TOEFL Reading Comprehension items.

From a practical standpoint, the results could help indicate the kind of information that might be provided by different categories of MC cloze items, if such items were to be considered for use in connection with the TOEFL. In particular, the data would show whether a MC cloze item type could be developed that relates specifically to reading comprehension more than to other aspects of English proficiency, in order to address the concerns about reading assessment raised earlier.

To address the above issues, the plan was to administer, in an operational testing situation, a test containing four sections: the three standard sections of the TOEFL and an experimental section containing 50 MC cloze items. A factor analysis of the basic TOEFL (i.e., the first three sections of this four-section test) would be

conducted to determine the degree to which the different parts of the TOEFL measured distinct skills. Then, the correlations of the various categories of MC cloze items with each other would be obtained. And, finally, correlational analysis would assess the relation of each category of MC cloze item to each part of the TOEFL. In this way, the study would determine whether the MC cloze items that are intended to tap a given skill were most highly related to the TOEFL items designed to measure that same skill.

The analyses just described were conducted for each of the nine language groups with the highest TOEFL candidate volumes. These groups included Arabic, Chinese, Farsi, French, Indonesian, Japanese, Korean, Spanish, and Thai. By performing separate analyses for each language group, the study provided data on interrelations among item types that would not be artifactually raised or lowered by conducting a single analysis with data from all language groups combined. Further, because separate analyses were conducted for each language group, it was possible to look for consistencies in results across groups, thus permitting conclusions about the generality of the observed results. And where differences were found, it would be possible to entertain hypotheses about the basis for the differences.

The research also dealt with a key statistical matter--estimation of parameters according to item response theory (IRT; Lord, 1980). IRT provides the basic method by which TOEFL test forms are equated. For this reason, it is essential that any item type considered for use in the TOEFL be amenable to IRT scaling. In IRT scaling, certain parameters must be determined for each item (see Results section). These parameters were estimated here in order to see whether MC cloze items are sufficiently amenable to IRT parameter estimation to warrant their eventual use in connection with the TOEFL.

## Method

### Subjects

The subjects of the study were 11,290 examinees who took the TOEFL at the November 1984 International administration at domestic test centers (i.e., test centers in the United States and Canada). The sample consisted of all examinees in the nine language groups containing the highest candidate volumes. The sample constituted 78% of the total candidate volume in domestic test centers.

The language groups represented and the number of examinees per group were: Arabic (1,833), Chinese (4,016), Farsi (476), French (341), Indonesian (640), Japanese (922), Korean (1,331), Spanish (1,293), and Thai (438). The sample thus consisted of several different language families: Indo-European (Spanish, French, and Farsi); Altaic (Japanese and Korean); Sino-Tibetan (Chinese); Austro-Tai (Indonesian and Thai); and Semitic (Arabic).

### Materials

The examinees were given a special test form consisting of the regular three sections of the TOEFL followed by a fourth section consisting of MC cloze items.

The TOEFL. The three sections of the TOEFL are (a) Listening Comprehension, (b) Structure and Written Expression, and (c) Vocabulary and Reading Comprehension. As outlined in the TOEFL Test and Score Manual (Educational Testing Service, 1987) the item types are as follows. In Listening Comprehension, the examinee hears spoken material (either single statements, short dialogues, or short monologues) and then hears questions about them, which he or she answers by selecting the correct answer choices in a test booklet. In the other two sections, all information is presented in written form. The Structure and Written Expression section consists of two item subtypes. In each Structure item, the examinee is given a portion of a sentence and, from the response alternatives, must choose the word or phrase that best completes the sentence. In each Written Expression item, the examinee is given a sentence in which words or phrases are underlined, and the examinee must indicate the underlined word or phrase that is ungrammatical. The third section consists of two basic item subtypes, Vocabulary and Reading Comprehension. In each Vocabulary item, a sentence is presented with a word or phrase underlined, and the examinee must choose the response alternative that is synonymous with the underlined word or phrase. In the Reading Comprehension part, segments of text are presented and, following each segment, several questions appear. The examinee must answer each question by selecting the correct response alternative.

The TOEFL form used in the study was that developed for use in

the November 1984 administration. Thus, although experimental cloze items were added to the test as a fourth section, the TOEFL used in the study was an operational test, prepared according to the usual procedures.

The cloze items. MC cloze items were developed for purposes of this study. To develop these items, six passages from previously used TOEFL forms were identified as representing an appropriate range of difficulty, according to statistical analyses of data from previous administrations.

Using these passages (and making some modifications as necessary), a total of 150 MC cloze items were developed, including the response options. A rational deletion procedure was used for identifying the words to omit, and items were developed that fit the categories in the classification scheme described below.<sup>2</sup>

On the basis of data from pretesting in three local universities (including students in the regular academic program and students in the intensive English program at each university to ensure a broad spectrum of proficiency), a final set of three passages was selected. These were passages for which the range of difficulty matched the range of difficulty encountered in the typical TOEFL. Then, to select the final 50 MC cloze items to include in the study, 25 items in the selected passages were eliminated by omitting the last few items per passage, including the sentences in which they appeared (except those sentences that were essential to the point of the paragraph). The final passages began with intact sentences and ended with one or two intact sentences. In the final set of 50 items, the number of words between deletions ranged from 4 to 15, with an average of 8.9.

A final step consisted of modifying either the stem or a response option in each of a few items, for either of the following reasons: (a) pretesting indicated that the item showed a low correlation with total MC cloze score, or (b) test-sensitivity review indicated the need for slight rewording.

The four sections of the test, including the cloze items in Section 4, are presented in Appendix A.

Classification scheme for cloze items. A classification scheme was devised for the MC cloze items. This scheme recognizes that the three skill areas of grammar, vocabulary, and reading comprehension are interrelated in certain respects and that reading comprehension is involved to some degree in all items. Thus, some items might primarily tap reading comprehension, as they require sensitivity to longer-range textual constraints, and secondarily require knowledge of vocabulary or knowledge of grammar. Other items might primarily tap

---

<sup>2</sup>The items, and the classification scheme, were developed by coauthors Oller and Butler.

vocabulary or grammar and secondarily involve reading comprehension, in that they require attention only to short-range constraints of semantics or syntax.

Thus, a four-category scheme was developed wherein the categories were defined as follows:

(a) Reading Comprehension/Grammar (RG) (9 items)

In this category, the task is one of understanding propositional information at an interclausal level, but answering the question also emphasizes knowledge of syntax (i.e., sequential arrangement and markers of such arrangements) rather than of lexicon.

Example (Item 3): A ballad is a folk song; however, a folk song is not a ballad [because, if, whether, unless] it tells a story.

(The brackets denote the place where a word has been deleted, and the words in the brackets are the response alternatives. In the actual test, the response alternatives were vertically aligned in a rectangular frame, as shown in Appendix A.)

(b) Reading Comprehension/Vocabulary (RV) (14 items)

In this category the problem is one of long-range constraints, but a lexical choice is required to solve it. The reader's task is basically one of understanding the text and getting the propositional information out of elements that may be some distance apart (usually across clause boundaries), yet a lexical choice is also required.

Example (Item 19): ...known as the Lost Sea. It is listed (in) the Guinness Book of World Records as the world's largest underground [water, body, lake, cave].

(Parentheses denote another place where a word has been deleted, the correct response here being the word "in.")

Note that with item types a and b, understanding of textual constraints is not necessarily required across a great physical distance; it does, however, require understanding of textual constraints at least across clause boundaries and taking account of inter- rather than intrapropositional relationships.

(c) Grammar/Reading Comprehension (GR) (15 items)

Here the source of item difficulty involves relatively short-range grammatical constraints--usually a few words on either side of the blank, or within a single grammatical phrase or clause. The item



primarily taps knowledge of surface syntax, and reading comprehension is involved primarily because the reader must understand within-clause propositional information.

Example (Item 2): It is generally understood that a ballad is a song that tells a story, (but) a folk song is not so [easy, easily, ease, easier] defined.

(d) Vocabulary/Reading Comprehension (VR) (12 items)

The primary aspect of this category is vocabulary (including idioms and collocations) although it also invokes reading comprehension to the extent of understanding the information presented within clause boundaries. The main source of difficulty, from the examinee's standpoint, is vocabulary--not grammar and not the understanding of long-range textual constraints.

Example (Item 9): In fact, there are folk songs for many occupations--railroading, [following, mustering, concentrating, herding] cattle, and so on.

Appendix A presents all of the cloze items used in the study, and Appendix B presents their category designations, as determined by a consensus of raters. (The rating procedure is described in the Results section.)

Procedure

The first three sections of the test--the basic TOEFL--were administered in the standard manner. A total of 33 minutes was allowed for completion of the fourth, cloze section. This amount of time apparently was more than adequate, as determined by observation of the students in pretesting and by reports from two test center supervisors that 33 minutes was about 5 minutes more than was needed.

So the examinees would not come unprepared for the cloze section, an explanatory flyer had been sent to them prior to the test. The flyer contained a sample passage with MC cloze items, and it briefly indicated the reasons for inclusion of the fourth section. The flyer is presented in Appendix C.

At the beginning of the test administration, the test supervisor reiterated the points made in the flyer. The supervisor explained (a) that the test would consist of four sections; (b) that the first three sections contained the standard TOEFL items and comprised the basis for the examinee's TOEFL score; and (c) that the fourth section contained a new type of item and that performance on this section would not count toward the TOEFL score but that the examinees should do their best. It was felt that, in an operational test setting, the examinees would be unlikely to risk answering any test items carelessly.

## Results

### Method of Analysis

Three sets of analyses were performed, each for the nine language groups separately. First, the basic TOEFL (excluding the cloze items) was factor analyzed to determine the degree to which the various parts of the TOEFL tap independent skills. Toward this end, a confirmatory factor analysis was employed, with an exploratory factor analysis performed as a preliminary step. Second, correlations among the cloze item categories were computed to determine the degree of independence among them. Finally, correlations were computed between the cloze item categories and the different parts of the TOEFL to determine whether cloze and TOEFL items that are intended to tap a particular skill relate more highly to each other than to items intended to measure other skills. Additional analyses examined the multiple regression of cloze performance on TOEFL performance and determined the degree to which cloze items are amenable to estimation of IRT parameters.

### Analyses of the TOEFL

The factor analyses of the TOEFL, though an initial step toward relating the MC cloze items to the TOEFL, were necessarily comprehensive. Thus, a substantial portion of the results section is devoted to discussion of these analyses.

For purposes of the present study, the second and third sections of the TOEFL were broken down into their two components, thus yielding five scores for analysis: (1) Listening Comprehension, (2) Structure, (3) Written Expression, (4) Vocabulary, and (5) Reading Comprehension. There were two reasons for subdividing the test in this fashion. First, the factor analysis by Swinton and Powers (1980) cited earlier had suggested that, depending on the language group studied, Vocabulary and Reading Comprehension may load on different factors; further, although the separation of Structure and Written Expression was not as clear, data from Swinton and Powers also indicated the merits of examining these two subsections separately. (Subparts of the Listening Comprehension section, however, appeared to be highly related to each other.) The second reason for breaking down the test in this fashion is that the original TOEFL consisted of five parts that were roughly comparable in format to the five parts listed above. In 1976, Structure and Written Expression were combined, as were Vocabulary and Reading Comprehension, based on correlational data from Pike (1979). Despite these correlational data, however, these five parts reflect somewhat different content and thus deserve separate examination in order to determine the relation of each to performance on the MC cloze test and its separate item categories.



The first results of interest, presented in Table 1, are the means and standard deviations of TOEFL scores for the nine language groups, as these data indicate the relative proficiency of the groups. Scores tended to be highest for the French and Spanish groups and lowest for the Japanese and Arabic groups, with scores of the other groups falling in between. These results are similar to those found in other research. Most pertinent are the studies by Alderman and Holland (1981) and Swinton and Powers (1980), who examined TOEFL data from seven language groups. (The same data base was used in both studies.) Five of the same language groups were used in the present study: Arabic, Chinese, Farsi, Japanese, and Spanish. As in the previous research, the highest-scoring language groups of the five were Spanish and Chinese. Also, in both cases, Arabic and Japanese examinees had scores that were similar and were at the bottom or near the bottom of the ranking. The one difference is that the Farsi group ranked at the bottom in the earlier research but in the middle of these five groups in the present research. This last result undoubtedly reflects a change due to the political situation in the early 1980s, such that Iranian students attending U.S. universities are now a more selective group.

Table 2 presents, for each language group, intercorrelations among the five parts of the TOEFL, the reliabilities of the scores, and the correlations corrected for unreliability. Most relevant for the present purposes are the corrected correlations, presented above the diagonal in each case. These correlations show that there is a relatively high degree of relationship among the last four parts of the test; the only part that appears to be somewhat separate from the others is Listening Comprehension. As will be seen below, factor analyses also suggest that the four nonlistening parts of the test tap highly related processes.

Exploratory factor analyses. The purpose of this set of analyses and the subsequent confirmatory factor analyses was to establish the factor structure of the TOEFL, so that the role of the MC cloze items in this structure could be established for each of the language groups. The exploratory factor analyses were performed as a preliminary step to conduction of the confirmatory factor analyses, which were intended to provide the primary basis for inferences about the test's structure.

For all factor analyses, each subsection of the test was divided into "parcels," with each parcel consisting of the total score for a group of items. There were three parcels of items for each of the five subsections of the test except Structure, for which there were two parcels. The items in each parcel were chosen so that, within each subsection, the parcels would be roughly equal in average difficulty and in distribution of item difficulties. These parcels served as the basic units of analysis.

Table 1  
Means and Standard Deviations of Raw Scores  
on the TOEFL for Each Language Group

Language Group	N	List. Comp. (Max.=50) <sup>a</sup>	Struc. (Max.=14)	Writ. Exp. (Max.=24)	Total		Vocab. (Max.=29)	Read. Comp. (Max.=29)	Vocab. and Read. Comp. (Max.=58)	Total TOEFL (Max.=146)
					Struc. and Writ. Exp. (Max.=38)					
Arabic	1833	Mean	8.47	15.11	23.58		14.73	13.36	28.09	81.73
		SD	2.76	4.93	7.15		5.99	5.35	10.63	25.08
Chinese	4016	Mean	9.71	16.78	26.49		18.06	18.04	36.10	95.19
		SD	2.37	3.96	5.84		5.24	5.05	9.62	21.69
Farsi	476	Mean	33.63	16.52	26.03		16.92	15.79	32.71	92.37
		SD	9.62	4.98	7.24		6.00	5.79	11.00	25.77
French	341	Mean	11.16	19.83	30.99		22.42	20.41	42.83	110.72
		SD	2.31	3.75	5.66		4.73	5.83	9.91	23.62
Indonesian	640	Mean	30.89	14.99	23.89		15.76	16.63	32.39	87.17
		SD	8.44	3.94	5.82		5.05	4.75	9.00	21.37
Japanese	922	Mean	29.25	15.86	25.07		15.81	15.09	30.91	85.23
		SD	9.68	4.52	6.85		5.83	5.86	11.00	23.39
Korean	1331	Mean	30.17	16.62	26.29		18.04	17.33	35.37	91.83
		SD	9.73	4.38	6.59		5.56	5.62	10.49	24.14
Spanish	1293	Mean	34.80	17.58	27.68		20.45	18.79	39.24	101.72
		SD	9.80	4.57	6.76		4.59	5.51	9.50	23.99
Thai	438	Mean	30.38	15.79	24.84		16.39	16.06	32.44	87.67
		SD	8.95	4.00	5.87		5.04	4.97	9.15	21.47

Each mean is the average number of correct items on the section indicated. For each of the four sections (Structure, Written Expression, Vocabulary, and Reading Comprehension), one item was inoperational; thus the maximum possible score on the test was 146 rather than 150.

<sup>a</sup>Max. = maximum score possible.

Table 2  
Correlations Among TOEFL Subscores and Reliabilities<sup>a</sup>

Language Group		List. Comp.	Struc.	Writ. Exp.	Vocab.	Read. Comp.
Arabic	List. Comp.	[.90]	.82	.80	.81	.80
	Struc.	.64	[.68]	.94	.92	.88
	Writ. Exp.	.69	.71	[.83]	.91	.89
	Vocab.	.71	.70	.77	[.85]	.92
	Read. Comp.	.68	.65	.72	.76	[.80]
Chinese	List. Comp.	[.88]	.77	.74	.80	.81
	Struc.	.58	[.64]	.99	.90	.86
	Writ. Exp.	.61	.69	[.75]	.88	.84
	Vocab.	.68	.65	.69	[.82]	.92
	Read. Comp.	.68	.62	.65	.75	[.80]
Farsi	List. Comp.	[.91]	.82	.79	.80	.84
	Struc.	.67	[.73]	.92	.88	.91
	Writ. Exp.	.69	.72	[.85]	.88	.85
	Vocab.	.71	.69	.75	[.86]	.88
	Read. Comp.	.74	.71	.72	.74	[.84]
French	List. Comp.	[.93]	.88	.83	.81	.88
	Struc.	.70	[.69]	.98	.99	.89
	Writ. Exp.	.72	.73	[.81]	.93	.87
	Vocab.	.71	.75	.76	[.83]	.89
	Read. Comp.	.79	.69	.73	.76	[.87]
Indonesian	List. Comp.	[.88]	.86	.80	.88	.83
	Struc.	.65	[.65]	.87	.87	.88
	Writ. Exp.	.64	.59	[.72]	.91	.88
	Vocab.	.74	.63	.69	[.80]	.88
	Read. Comp.	.68	.62	.65	.68	[.76]
Japanese	List. Comp.	[.90]	.79	.81	.81	.84
	Struc.	.63	[.71]	.98	.92	.94
	Writ. Exp.	.69	.74	[.80]	.91	.89
	Vocab.	.71	.71	.75	[.85]	.92
	Read. Comp.	.73	.73	.73	.77	[.84]
Korean	List. Comp.	[.91]	.68	.73	.73	.76
	Struc.	.55	[.72]	.97	.89	.88
	Writ. Exp.	.62	.73	[.79]	.90	.87
	Vocab.	.64	.70	.74	[.85]	.91
	Read. Comp.	.66	.68	.71	.76	[.84]

Table 2 (continued)

Language Group		List. Comp.	Struc.	Writ. Exp.	Vocab.	Read. Comp.
Spanish	List. Comp.	[.92]	.83	.86	.83	.79
	Struc.	.67	[.70]	.97	.91	.88
	Writ. Exp.	.75	.74	[.83]	.90	.85
	Vocab.	.71	.68	.73	[.79]	.94
	Read. Comp.	.70	.68	.72	.77	[.84]
Thai	List. Comp.	[.89]	.78	.70	.72	.79
	Struc.	.58	[.63]	.94	.86	.89
	Writ. Exp.	.57	.64	[.73]	.85	.89
	Vocab.	.60	.61	.65	[.80]	.85
	Read. Comp.	.66	.62	.67	.67	[.78]

<sup>a</sup> Coefficient alpha reliabilities are in brackets. Raw correlations appear below the diagonal, and correlations corrected for attenuation appear above the diagonal.

The first step was a principal components analysis. The initial results of interest were the eigenvalues resulting from this analysis, which are shown in Table 3. For every language group, the first eigenvalue was very large, ranging from 7.10 (accounting for 51% of the total variance) to 8.73 (62% of the total variance), with a median across language groups of 8.25 (59% of the total variance). The second eigenvalues were generally much lower, ranging from .82 to 1.16 (median = .84). The first two eigenvalues together accounted for 58% to 68% of the variance (median = 66%). Other eigenvalues were below .76. Although the benefit derived in going from one factor to two apparently was not substantial, as most of the second eigenvalues were below 1, the latter eigenvalues were close enough to 1 to warrant entertaining the possibility of a two-factor solution.

Varimax orthogonal rotations were performed for (a) a two-factor solution, for reasons just mentioned, (b) a three-factor solution, since the earlier Swinton and Powers (1980) study had suggested the presence of three factors, and (c) a five-factor solution, because it was of interest to see the extent to which the test's factor structure would be linked to the test's five subsections. Output from the Varimax solutions are presented in Appendix D.

Although the data from the Varimax rotations were cleaner for some language groups than others, with some indication of a clear distinction among factors for certain groups, generally the solutions were not clear-cut. One reason for this is that the Varimax technique tends to capitalize on quite small differences in correlations among parcels, with only a .01 or .02 point difference in correlations often determining whether a parcel loads on one factor rather than another. More importantly, a Varimax solution may yield what appears to be several factors, but since it is an orthogonal solution it cannot show the extent to which the parts of the test are correlated with each other. Thus, determining the number of factors that meaningfully distinguish the skills measured by the TOEFL requires a method that permits assessment of the degree of relationship among parts of the test. The confirmatory factor analyses, discussed in the next section, provide such information, as they permit examination of the correlations among predefined factors.<sup>3</sup>

Confirmatory factor analyses. A set of confirmatory factor analyses for each language group (Joreskog, 1971) was conducted, using the maximum likelihood estimation procedure of LISREL VI (Joreskog & Sorbom, 1981, 1983). Parallel analyses were conducted on the correlation and variance-covariance matrices. The results of these

---

<sup>3</sup>Although an oblique rotation could also have been conducted to obtain information relevant to this issue within an exploratory factor-analytic framework, it was decided to proceed directly to the confirmatory factor analyses.

Table 3

First Six Eigenvalues from Principal Components Factor Analysis  
of the TOEFL and Cumulative Percentage of Total Variance  
Accounted for, by Language Group

<u>Language Group</u>	<u>Eigenvalue</u>	<u>Percentage of Total Variance</u>
Arabic	8.25	59
	.84	65
	.66	70
	.54	74
	.52	77
	.47	81
Chinese	7.55	54
	.98	61
	.75	66
	.56	70
	.55	74
	.53	78
Farsi	8.48	61
	.83	67
	.62	71
	.61	75
	.52	79
	.43	82
French	8.73	62
	.82	68
	.65	73
	.57	77
	.48	80
	.44	84
Indonesian	7.36	53
	.82	58
	.76	64
	.69	69
	.58	73
	.57	77
Japanese	8.48	61
	.86	67
	.63	71
	.54	75
	.48	79
	.48	79

Table 3 (continued)

<u>Language Group</u>	<u>Eigenvalue</u>	<u>Percentage of Total Variance</u>
Korean	8.05	57
	1.16	66
	.68	71
	.56	75
	.51	78
	.44	81
Spanish	8.40	60
	.82	66
	.69	71
	.54	75
	.49	78
	.48	81
Thai	7.10	51
	1.08	58
	.76	64
	.71	69
	.66	74
	.64	78

two sets of analyses were basically the same. It was decided to focus on the analyses of the correlation matrices because of the relative ease of interpreting their results.

The degree of fit of different factor solutions (i.e., solutions differing in number of factors used) was assessed by two indices: the goodness of fit index, and the root mean square residual.<sup>4</sup> The goodness of fit index is a value that approaches unity as its optimal limit; the closer the index is to unity for a given solution, the greater is the likelihood that the number of factors in that solution is sufficient to account for the data. The root mean square residual is a number that is generally near zero when the fit of the model is good. In effect, it reflects the average portion of the observed correlations that is not attributed to the factors that have been defined. When two possible solutions are attempted, such as a two-factor solution and a three-factor solution, if these indices are not substantially different for the two solutions, the solution involving the smaller number of factors is generally preferred.

For the present purposes, four solutions were attempted. First, because of the design structure of the test, and the goal of relating MC cloze performance to each part of the TOEFL, a five-factor solution was tried, in which the five factors were Listening Comprehension, Structure, Written Expression, Vocabulary, and Reading Comprehension. A second solution was one involving three factors, in which the factors reflected the three sections of the TOEFL: (a) Listening Comprehension, (b) Structure and Written Expression, and (c) Vocabulary and Reading Comprehension. A third solution was one involving two factors, which related to (a) Listening Comprehension, and (b) all other subtests. The basis for this last analysis lay in the two-factor Varimax results (Appendix D); to the extent that two factors could be identified, the highest loadings on the first factor were associated with Listening Comprehension and those on the second factor were associated with the remainder of the test. A fourth analysis was a one-factor solution, which was conducted to permit comparison with the other solutions with respect to goodness of fit. The factor loadings for each of the four solutions are presented in Appendix E.

The goodness of fit index and root mean square residual for each solution and each language group are presented in Table 4. Whereas the goodness of fit indices for the one-factor solution generally ranged from the low to high .80s (median across language groups = .86), the indices for the two-, three-, and five-factor solutions ranged from the low to high .90s (medians for the three solutions =

---

<sup>4</sup>A maximum likelihood ratio chi-square index is sometimes examined in this context, but it was not used here because it is sample-size dependent, and the present study used a very large sample. With samples of sufficient size, even trivial departures from fit will yield a significant chi-square.



Table 4

Indices from Confirmatory Factor Analysis of the TOEFL  
for One-, Two-, Three-, and Five-Factor Solutions<sup>a</sup>

Language Group	One-Factor Solution			Two-Factor Solution			Three-Factor Solution			Five-Factor Solution		
	Goodness of Fit Index	Root Mean Square Residual		Goodness of Fit Index	Root Mean Square Residual		Goodness of Fit Index	Root Mean Square Residual		Goodness of Fit Index	Root Mean Square Residual	
Arabic	.89	.04		.97	.02		.98	.02		.99	.02	
Chinese	.89	.04		.96	.03		.99	.02		.98	.02	
22 Farsi	.86	.04		.94	.03		.95	.03		.98	.02	
French	.83	.04		.92	.03		.93	.03		.96	.02	
Indonesian	.92	.04		.96	.03		.96	.03		.98	.02	
Japanese	.87	.04		.98	.03		.97	.02		.98	.02	
Korean	.81	.06		.94	.03		.96	.03		.98	.03	
Spanish	.86	.04		.95	.03		.97	.02		.99	.01	
Thai	.85	.05		.94	.04		.95	.03		.97	.03	

<sup>a</sup>In the two-factor solution, the two factors are: (a) Listening Comprehension, and (b) all other parts of the test; in the three-factor solution, the factors are: (a) Listening Comprehension, (b) Structure and Written Expression, and (c) Vocabulary and Reading Comprehension; in the five-factor solution, the factors are: (a) Listening Comprehension, (b) Structure, (c) Written Expression, (d) Vocabulary and, (e) Reading Comprehension.

.95, .96, .98). Thus, the difference between one- and two-factor solutions in goodness of fit was reasonably large. Hence, a two-factor solution apparently provides a better fit to the data than a one-factor solution. The results differ slightly between the two- and three-factor solutions, and between the three- and five-factor solutions. But the differences are so small, and the indices so near their limits with the two-factor solution, that two factors appear sufficient to account for performance on the TOEFL.

Another aspect of the data that supports the above conclusion is the correlations among factors in the two-, three-, five-factor solutions, which appear in Tables 5 and 6. In the two-factor solution, the correlation between the Listening Comprehension factor and the nonlistening factor ranged from .76 to .89 across language groups, with a median of .84. In the three-factor solution, the correlations involving the Listening Comprehension factor ranged from .71 to .89, with a median of .83, whereas the correlations between the two non-listening factors ranged from .89 to .96, with a median of .94. In the five-factor solution, the correlations involving the Listening Comprehension factor ranged from .68 to .88, with a median of .81, whereas the correlations among the other four factors ranged from .83 to .99, with a median of .90. Thus, a reasonable conclusion is that the Listening Comprehension factor is less highly correlated with the other factors than the latter are with each other (although the correlations between the Listening Comprehension factor and the others are still relatively high).

Regarding the correlations among the nonlistening factors in the five-factor solutions, there did not appear to be any consistent pattern across language groups except for the fact that the correlation between factors 2 (Structure) and 3 (Written Expression) was generally the highest in the cluster, with a median across language groups of .97. However, factors 4 and 5 (Vocabulary and Reading Comprehension) did not show consistently higher correlations with each other than with factors 2 and 3. For example, the Vocabulary factor showed median correlations of .90, .91, and .90, respectively, with Structure, Written Expression, and Reading Comprehension. So it must be concluded that these four factors do not divide into clearly identifiable subsets.

In sum, two factors at most appear sufficient to account for performance on the TOEFL. Even a one-factor solution can account for a sizable portion of the variance in performance, although there seems to be some merit in regarding the Listening Comprehension factor as somewhat distinct from factors associated with the other parts of the test. Perhaps most important for purposes of the present study, there seems to be little empirical indication that the nonlistening parts of the test--Structure, Written Expression, Vocabulary, and Reading Comprehension--measure separate aspects of English proficiency.

Table 5  
Correlations among Factors in Two- and Three-Factor  
Confirmatory Factor Analyses of the TOEFL

Language Group	Two-Factor Analysis	Three-Factor Analysis		
	List. Comp.		List. Comp.	Struc. & Writ. Exp.
Arabic				
Non-list. <sup>a</sup>	.84	Struc. & Writ. Exp.	.83	
		Vocab. & Read. Comp.	.83	.94
Chinese				
Non-list.	.82	Struc. & Writ. Exp.	.76	
		Vocab. & Read. Comp.	.82	.89
Farsi				
Non-list.	.86	Struc. & Writ. Exp.	.82	
		Vocab. & Read. Comp.	.85	.94
French				
Non-list.	.88	Struc. & Writ. Exp.	.85	
		Vocab. & Read. Comp.	.87	.94
Indonesian				
Non-list.	.89	Struc. & Writ. Exp.	.85	
		Vocab. & Read. Comp.	.89	.96
Japanese				
Non-list.	.84	Struc. & Writ. Exp.	.80	
		Vocab. & Read. Comp.	.85	.94
Korean				
Non-list.	.76	Struc. & Writ. Exp.	.71	
		Vocab. & Read. Comp.	.76	.92
Spanish				
Non-list.	.86	Struc. & Writ. Exp.	.86	
		Vocab. & Read. Comp.	.82	.90
Thai				
Non-list.	.78	Struc. & Writ. Exp.	.73	
		Vocab. & Read. Comp.	.79	.93

<sup>a</sup>"Non-list." represents the factor defined by the four nonlistening parts of the TOEFL: Structure, Written Expression, Vocabulary, and Reading Comprehension.

Table 6

Correlations among Factors in Five-Factor  
Confirmatory Factor Analysis of the TOEFL

Language Group	Factor			
	List. Comp.	Struc.	Writ. Exp.	Vocab.
Arabic				
Struc.	.82			
Writ. Exp.	.81	.93		
Vocab.	.81	.91	.91	
Read. Comp.	.80	.87	.88	.91
Chinese				
Struc.	.77			
Writ. Exp.	.75	.98		
Vocab.	.80	.89	.88	
Read. Comp.	.81	.85	.83	.92
Farsi				
Struc.	.85			
Writ. Exp.	.79	.94		
Vocab.	.80	.90	.89	
Read. Comp.	.84	.94	.86	.87
French				
Struc.	.84			
Writ. Exp.	.84	.94		
Vocab.	.81	.94	.94	
Read. Comp.	.87	.85	.88	.89
Indonesian				
Struc.	.85			
Writ. Exp.	.79	.87		
Vocab.	.88	.86	.91	
Read. Comp.	.82	.86	.87	.86
Japanese				
Struc.	.78			
Writ. Exp.	.80	.97		
Vocab.	.81	.90	.91	
Read. Comp.	.84	.92	.89	.91

Table 6 (continued)

Language Group	Factor			
	List. Comp.	Struc.	Writ. Exp.	Vocab.
Korean				
Struc.	.68			
Writ. Exp.	.72	.97		
Vocab.	.73	.89	.91	
Read. Comp.	.75	.87	.86	.90
Spanish				
Struc.	.84			
Writ. Exp.	.85	.97		
Vocab.	.83	.91	.90	
Read. Comp.	.79	.88	.85	.93
Thai				
Struc.	.85			
Writ. Exp.	.69	.99		
Vocab.	.72	.95	.84	
Read. Comp.	.78	.97	.87	.84

## Analyses of MC Cloze Items

Content analysis. As indicated above, the MC cloze items were subdivided into four different categories: Reading Comprehension/Grammar (RG), Reading Comprehension/Vocabulary (RV), Grammar/Reading Comprehension (GR), and Vocabulary/Reading Comprehension (VR). In addition to the two persons who developed the classification scheme, two other persons with expertise in ESL were given descriptions of the four categories and the three passages with the 50 MC cloze items and were asked to indicate the category in which each item belonged.<sup>5</sup> Thus, there were four persons evaluating the categories of the items. Although the first two persons developed the classification scheme, they had worked independently; thus they were treated as independent judges for the present purpose.

The four judges were asked to indicate the category in which each item should be placed. They were told: "You may feel that some items clearly fall into the categories you have indicated, whereas you may be less sure of others. For each item, we would like you to indicate your degree of certainty in this regard on a four-point scale: Very certain (4), Somewhat certain (3), Somewhat uncertain (2), and Very uncertain (1). For those items you do not rate as Very certain, please indicate the other category (or categories) in which the item could also be placed, in your judgment."

Of the 50 items, 14 were placed in the same category by all four judges and 30 were placed in the same category by three of the four judges (and for 15 of these 30 items, the fourth judge's secondary rating was the same as the other judges' primary rating). Of the remaining 6 items, 3 were rated in one category by two judges and another category by two others, while the other 3 were placed in one category by two judges and in each of two other categories by the other judges.

For purposes of the analyses to follow, each of the 50 items was placed into a category based on the judges' ratings. Where three or four judges picked the same category (44 of the 50 items), the item was assigned to that category. Where two judges picked one category but the other two judges were split between two other categories (3 items), the item was assigned to the first category. Where two judges picked one category and the other two picked another category (3 items), the item was assigned to the category of the judges who attributed greater confidence to their ratings. Appendix B presents the categories in which the items were placed.

Empirical analysis of item classification. Table 7 presents the means and standard deviations of the MC cloze scores. The different

---

<sup>5</sup>The authors gratefully acknowledge the contribution of Lyle Bachman of the University of Illinois and Marianne Celce-Murcia of the University of California at Los Angeles in this phase of the project.

Table 7

## Means and Standard Deviations of Cloze Scores

Language Group		Cloze Item Category				Total Cloze Score (Max.=50)
		RG (Max.=9)	RV (Max.=14)	GR (Max.=15)	VR (Max.=12)	
Arabic	Mean	4.44	5.28	9.33	6.47	25.52
	<u>SD</u>	1.93	2.68	3.15	2.44	8.53
Chinese	Mean	5.16	6.70	10.68	7.62	30.16
	<u>SD</u>	1.88	2.64	2.46	2.33	7.56
Farsi	Mean	5.08	6.14	10.16	7.18	28.55
	<u>SD</u>	2.03	2.97	2.99	2.68	9.19
French	Mean	6.07	9.63	12.76	9.09	37.55
	<u>SD</u>	1.84	2.95	2.21	2.20	7.92
Indonesian	Mean	4.44	5.76	10.00	6.65	26.86
	<u>SD</u>	1.81	2.23	2.35	2.25	6.76
Japanese	Mean	4.65	6.44	10.05	6.96	28.10
	<u>SD</u>	1.99	2.77	2.70	2.49	8.22
Korean	Mean	4.46	6.43	10.03	7.41	28.33
	<u>SD</u>	1.97	2.79	2.85	2.45	8.45
Spanish	Mean	5.59	8.68	11.75	8.41	34.44
	<u>SD</u>	1.94	2.84	2.68	2.28	8.32
Thai	Mean	4.67	5.84	10.20	7.10	27.81
	<u>SD</u>	1.80	2.40	2.44	2.28	7.15

language groups tended to rank order in the same manner as they did for TOEFL performance, with the French group showing the highest scores and the Arabic group, the lowest.

The total score for each MC cloze category was computed, and correlations among these scores were calculated. Also, reliabilities of the four categories were computed. These data are presented in Table 8. Although the correlations were generally in the .50s and .60s, the reliabilities of the scores for the different MC cloze categories were typically within the same range. Hence, as measured against the reliabilities, these correlations may be regarded as quite high.

Variation in reliability among the categories could be due to differences among the MC cloze subscores in number of items contributing to them--9, 14, 15, and 12, respectively, for RG, RV, GR, and VR. Thus, to permit comparison of the reliability coefficients, the Spearman-Brown formula was applied in order to yield estimates of the reliabilities that would have been observed if each MC cloze category had contained 15 items. The results are presented in Table 8 under the heading "corrected reliability." With the corrections it can be seen that the reliabilities of the four MC cloze item categories are relatively comparable. (Note that, overall, the reliabilities appear to be slightly higher for the European and Semitic languages--French, Spanish, and Farsi--than for the others).

In sum, the key finding of the analyses of MC cloze items is that the correlations do not provide a clear basis on which to conclude that the four categories are measuring distinct processes.

#### The Relation of MC Cloze Items to the TOEFL

Correlations among cloze and TOEFL scores. The main objective of the study was to determine how the MC cloze items relate to the different parts of the TOEFL, for each language group. In this regard, correlations were computed among the MC cloze and TOEFL scores, the former consisting of the scores for each MC cloze item category and total MC cloze score, and the latter consisting of the scores for the five parts of the TOEFL and the total TOEFL score. The correlations are presented in Table 9, and the correlations corrected for unreliability are presented in Table 10. The discussion to follow is based on the data in the latter table.

It is of value, first, to examine the correlations between the total MC cloze score and the TOEFL part and total scores. The correlation between total MC cloze score and total TOEFL score ranged from .86 to .94 across language groups (median = .89). Thus, the relationship between these two types of test was quite high, indicating a high degree of overlap in the processes measured by them.



Table 8  
Correlations among Cloze Scores and Reliabilities

Language Group	RG	RV	GR	VR	Reliability	Corrected Reliability <sup>a</sup>	Language Group	RG	RV	GR	VR	Reliability	Corrected Reliability
Arabic	RG						Japanese	RG					
	RV	.51			.48	.61		RV	.54			.53	.65
	GR	.60	.59		.51	.53		GR	.59	.55		.64	.67
	VR	.55	.60	.67	.55	.60		VR	.55	.58	.62	.62	.67
TOTAL CLOZE		.77	.82	.88	.85	.87	TOTAL CLOZE		.78	.83	.85	.86	
Chinese	RG						Korean	RG					
	RV	.48			.48	.61		RV	.54			.52	.64
	GR	.51	.56		.60	.62		GR	.58	.62		.66	.68
	VR	.51	.57	.59	.58	.63		VR	.56	.64	.65	.61	.66
TOTAL CLOZE		.74	.83	.83	.84		TOTAL CLOZE		.77	.85	.86	.87	
Farsi	RG						Spanish	RG					
	RV	.57			.56	.68		RV	.60			.55	.67
	GR	.61	.66		.72	.71		GR	.64	.64		.69	.70
	VR	.59	.68	.75	.70	.74		VR	.61	.63	.69	.73	.73
TOTAL CLOZE		.78	.86	.89	.89		TOTAL CLOZE		.81	.86	.88	.89	.71
French	RG						Thai	RG					
	RV	.65			.54	.66		RV	.47			.40	.53
	GR	.58	.67		.74	.75		GR	.53	.49		.53	.55
	VR	.62	.70	.66	.68	.68		VR	.50	.55	.57	.57	.61
TOTAL CLOZE		.81	.90	.85	.89	.72	TOTAL CLOZE		.75	.80	.82	.81	
Indonesian	RG												
	RV	.46			.42	.51							
	GR	.49	.49		.51	.53							
	VR	.47	.50	.48	.52	.52							
TOTAL CLOZE		.77	.79	.80	.79	.58							

<sup>a</sup> Reliability that would be observed if the category of cloze item contained 15 items, as estimated by Spearman-Brown formula.

Table 9  
Correlations of Cloze and TOEFL Scores

Language Group	Cloze Category	TOEFL Subscore					Total TOEFL Score
		List. Comp.	Struc.	Writ. Exp.	Vocab.	Read. Comp.	
Arabic	RG	.49	.52	.56	.57	.53	.60
	RV	.50	.50	.55	.64	.63	.64
	GR	.57	.61	.65	.64	.62	.69
	VR	.54	.54	.59	.62	.61	.66
	Total Cloze	.63	.66	.71	.74	.72	.78
Chinese	RG	.48	.51	.53	.51	.51	.58
	RV	.54	.51	.52	.60	.60	.65
	GR	.55	.55	.58	.60	.59	.66
	VR	.55	.54	.57	.63	.61	.68
	Total Cloze	.66	.65	.68	.73	.72	.80
Farsi	RG	.57	.57	.56	.60	.61	.66
	RV	.56	.58	.63	.65	.70	.70
	GR	.63	.66	.68	.66	.65	.74
	VR	.62	.62	.66	.68	.67	.74
	Total Cloze	.69	.71	.74	.75	.77	.83
French	RG	.59	.60	.68	.67	.61	.69
	RV	.64	.66	.65	.74	.72	.76
	GR	.63	.67	.73	.70	.67	.74
	VR	.66	.65	.69	.73	.67	.76
	Total Cloze	.73	.75	.79	.83	.78	.86
Indonesian	RG	.55	.50	.52	.56	.51	.62
	RV	.51	.44	.45	.57	.52	.59
	GR	.59	.55	.53	.59	.54	.66
	VR	.58	.50	.53	.63	.51	.65
	Total Cloze	.71	.63	.65	.75	.67	.80
Japanese	RG	.61	.61	.62	.63	.61	.69
	RV	.57	.55	.57	.65	.63	.67
	GR	.60	.65	.68	.69	.65	.73
	VR	.60	.61	.62	.68	.64	.71
	Total Cloze	.72	.73	.75	.81	.77	.85
Korean	RG	.53	.54	.57	.56	.58	.64
	RV	.55	.52	.56	.61	.66	.67
	GR	.57	.57	.61	.60	.62	.69
	VR	.54	.53	.58	.61	.63	.67
	Total Cloze	.66	.64	.69	.71	.74	.80

Table 9 (continued)

Language Group	Cloze Category	TOEFL Subscore					Total TOEFL Score
		List. Comp.	Struc.	Writ. Exp.	Vocab.	Read. Comp.	
Spanish	RG	.56	.61	.67	.62	.63	.69
	RV	.52	.56	.59	.67	.68	.67
	GR	.60	.64	.70	.68	.66	.73
	VR	.60	.60	.65	.67	.67	.72
	Total Cloze	.67	.70	.76	.77	.77	.82
Thai	RG	.46	.49	.55	.45	.54	.58
	RV	.50	.44	.50	.51	.60	.61
	GR	.55	.51	.58	.48	.54	.63
	VR	.50	.50	.54	.55	.52	.61
	Total Cloze	.63	.60	.68	.63	.69	.76

Table 10

## Correlations of Cloze and TOEFL Scores Corrected for Unreliability

Language Group	Cloze Category	TOEFL Subscore					Total TOEFL Score
		List. Comp.	Struc.	Writ. Exp.	Vocab.	Read. Comp.	
Arabic	RG	.74	<u>.91<sup>a</sup></u>	.88	.88	.84	.88
	RV	.67	<u>.77</u>	.76	.87	<u>.89</u>	.83
	GR	.70	<u>.86</u>	.84	.82	<u>.81</u>	.84
	VR	.73	<u>.84</u>	.83	.86	<u>.87</u>	.86
	Total Cloze	.72	.85	.84	.87	<u>.86</u>	.86
Chinese	RG	.74	<u>.91</u>	.88	.81	.82	.86
	RV	.74	<u>.82</u>	.77	.86	<u>.87</u>	.86
	GR	.75	<u>.89</u>	.87	.85	<u>.86</u>	.88
	VR	.77	<u>.88</u>	.87	<u>.92</u>	.90	.91
	Total Cloze	.76	.89	.86	<u>.88</u>	.88	.90
Farsi	RG	.80	.89	.82	.87	<u>.89</u>	.90
	RV	.70	.81	.81	.84	<u>.91</u>	.85
	GR	.78	<u>.91</u>	.87	.84	<u>.84</u>	.89
	VR	.78	<u>.87</u>	.85	.88	<u>.88</u>	.90
	Total Cloze	.77	.88	.85	.86	<u>.89</u>	.89
French	RG	.83	.98	<u>.99</u>	.99	.89	.96
	RV	.77	.93	<u>.83</u>	<u>.94</u>	.90	.89
	GR	.79	.97	<u>.98</u>	<u>.93</u>	.87	.92
	VR	.85	.95	<u>.94</u>	<u>.99</u>	.88	.95
	Total Cloze	.81	.96	.94	<u>.96</u>	.89	.93
Indonesian	RG	.91	<u>.95</u>	.94	.95	.91	.98
	RV	.76	<u>.76</u>	.75	<u>.89</u>	.84	.85
	GR	.88	<u>.94</u>	.86	<u>.92</u>	.87	.94
	VR	.85	<u>.85</u>	.87	<u>.97</u>	.80	.92
	Total Cloze	.86	.88	.86	<u>.94</u>	.86	.93
Japanese	RG	.88	<u>.99</u>	.95	.95	.92	.98
	RV	.74	<u>.81</u>	.79	<u>.88</u>	.85	.85
	GR	.78	<u>.95</u>	.95	<u>.93</u>	.89	.93
	VR	.80	<u>.92</u>	.88	<u>.94</u>	.89	.92
	Total Cloze	.81	.93	.91	<u>.95</u>	.91	.94
Korean	RG	.78	.89	<u>.89</u>	.85	.87	.91
	RV	.71	.75	<u>.77</u>	.82	<u>.89</u>	.85
	GR	.73	.81	<u>.83</u>	.79	<u>.81</u>	.85
	VR	.73	.80	<u>.84</u>	.84	<u>.87</u>	.87
	Total Cloze	.74	.82	.84	.83	<u>.87</u>	.88

Table 10 (continued)

Language Group	Cloze Category	TOEFL Subscore					Total TOEFL Score
		List. Comp.	Struc.	Writ. Exp.	Vocab.	Read. Comp.	
Spanish	RG	.79	.99	<u>.99</u>	.94	.93	.95
	RV	.65	.81	<u>.77</u>	.91	<u>.89</u>	.82
	GR	.74	.89	<u>.90</u>	.90	<u>.84</u>	.88
	VR	.77	.87	<u>.88</u>	<u>.92</u>	.89	.90
	Total Cloze	.74	.89	.89	<u>.93</u>	.90	.89
Thai	RG	.77	.99	<u>.99</u>	.81	.97	.95
	RV	.72	.76	<u>.81</u>	.79	<u>.94</u>	.86
	GR	.77	.86	<u>.90</u>	.72	<u>.81</u>	.87
	VR	.70	.84	<u>.84</u>	.82	.78	.84
	Total Cloze	.74	.85	<u>.88</u>	.78	.86	.87

<sup>a</sup>Highest correlation between cloze and TOEFL part scores in each row is underlined determined by correlations expressed to three decimal places in cases of apparent ties.

Correlations of the total MC cloze score with TOEFL Vocabulary or Reading Comprehension scores tended to be the highest of the correlations with the five parts of the TOEFL (seven out of nine cases). However, the differences in correlations for the four nonlistening parts of the TOEFL were not substantial; only the correlations involving Listening Comprehension were markedly lower than the others.

A principal issue in the study was to determine how each of the MC cloze item categories would relate to the various parts of the TOEFL. Evidence bearing on this issue is found in the corrected correlations between scores on the different MC cloze and TOEFL item types. Perhaps the most salient aspect of the data is that the correlations of the MC cloze scores with TOEFL Listening Comprehension were consistently lower (median across cloze item categories and language groups = .77) than their correlations with the other parts of the TOEFL (median across TOEFL parts, MC cloze categories, and language groups = .88). Hence, the MC cloze items apparently measure processes more closely related to those tapped by TOEFL Structure, Written Expression, Vocabulary and Reading Comprehension than by TOEFL Listening Comprehension. This finding is consistent with other results discussed above.

Regarding the four nonlistening parts of the TOEFL, the RG and GR categories of MC cloze items correlated slightly more highly with the Structure and Written Expression parts of the TOEFL, while the RV and VR categories correlated slightly more highly with the Vocabulary and Reading Comprehension parts. This can be seen in the fact that the highest correlations for the categories RG and GR (underlined in Table 10) were generally associated with Structure or Written Expression (17 of 18 cases), and the highest correlations for the categories RV and VR were generally associated with Vocabulary or Reading Comprehension (17 of 18 cases).

While the consistency of this pattern is provocative and worthy of further investigation, the differences among correlations were not very great in a practical sense. The following calculations help demonstrate the point. The average of all 36 correlations of RG and GR scores with the TOEFL Structure and Written Expression sections was .91; and the average of all 36 correlations of RG and GR scores with the Vocabulary and Reading Comprehension sections was .87. Each of these two average correlations was squared, and the difference between the squares was computed. The result reflected the percentage difference in degree to which the RG-GR cloze items shared variance with (a) the TOEFL Structure and Written Expression subsections, and (b) the TOEFL Vocabulary and Reading Comprehension subsections. On average, the RG-GR items shared only 7.5% more variance with the first of these two pairs of subsections than with the second. A similar analysis was done for the RV and VR cloze items. Average correlations of .83 and .88 were observed, indicating that the RV-VR items shared only 8.3% less variance with the TOEFL Structure and Written Expression subsections than with the Vocabulary and Reading Comprehension

subsections. Therefore, the differences in correlations were not substantial enough to conclude that one can identify subsets of MC cloze items that are clearly more related to some nonaural TOEFL item types than others. On the other hand, it must be remembered that the nonlistening subsections of the TOEFL were highly interrelated. Thus, it would be inappropriate to expect that a given classification of MC cloze items could correlate in a substantially different manner with one subsection than with the others.

Reliabilities of cloze and TOEFL scores. In assessing the merits of the MC cloze format as a possible addition to the item formats currently used in the TOEFL, it is important to determine how the cloze and TOEFL nonlistening items compare in reliability. Toward this end, it is necessary first to adjust for differences in the numbers of MC cloze and TOEFL nonlistening items (50 of 96, respectively). The Spearman-Brown formula was applied to the reliabilities of the MC cloze scores, so that the reliability of the MC cloze and TOEFL scores could be compared as if each had contained 96 items, as in the nonlistening portion of the TOEFL.

The adjusted MC cloze and TOEFL nonlistening reliabilities, respectively, were: Arabic .93 and .94; Chinese .91 and .93; Farsi .94 and .95; French .94 and .94; Indonesian .88 and .91; Japanese .92 and .94; Korean .93 and .94; Spanish .94 and .94; and Thai .89 and .92. Thus, the estimated reliability of the MC cloze items was quite close to that of the nonlistening portion of the TOEFL.

Multiple regression analysis. An additional question of interest concerned the extent to which performance on the MC cloze items contributed unique information about an examinee's language proficiency beyond that already furnished by the TOEFL. To address this question, multiple regression analyses were run for each language group, in which the dependent variable was total score on the MC cloze items and the independent variables were the five subparts of the TOEFL: Listening Comprehension, Structure, Written Expression, Vocabulary, and Reading Comprehension. (This analysis could not be run for the Thai group because the collinearities, or correlations, among certain parts of the TOEFL were too close to unity.)

The resulting multiple correlations were quite high: Arabic, .89; Chinese, .92; Farsi, .92; French, .97; Indonesian, .96; Japanese, .97; Korean, .89; and Spanish, .95. Thus, cloze performance can be nearly completely predicted by scores on the TOEFL. Apparently, then, the MC cloze scores provide little information beyond that provided by the TOEFL scores.

#### IRT Parameter Estimation, Test of Unidimensionality, and Equating of MC Cloze Items

This section describes the results of IRT analyses of the 50 MC cloze items, for all language groups combined, including an evaluation

of the goodness of fit of the items to the IRT model, Bejar's (1980) test of the unidimensionality of a set of items, and the reasonableness of equating based on the cloze items.

TOEFL tests are equated by means of the three-parameter IRT model (Lord, 1980). According to item response theory, the probability of answering an item correctly is a mathematical function of ability level. The function used here, called the "logistic" function, consists of one parameter for each examinee--ability level--and three parameters describing each item. The item parameters represent the item's discrimination power (the "a parameter"), the item's difficulty level (the "b parameter"), and the probability that low-ability examinees will guess the correct answer (the "c parameter").

The function that relates the probability of correctly answering a given item for an examinee of a given ability is an ogive (Figure 1 shows examples for six different items). The a parameter is reflected in the slope of the curve at its inflection point. The steeper the slope, the greater the item's capacity to discriminate among examinees within a narrow range of ability. The b parameter is reflected in the position of the curve on the horizontal axis; the further to the right the curve is situated, the more difficult the item. The c parameter, or "guessing" parameter, is the lower asymptote of the curve, which indicates the likelihood that a person with very low ability will answer the item correctly.

The MC cloze items were scaled (i.e., IRT parameters were estimated) by using data from TOEFL Section 2 (Structure and Written Expression) and Section 3 (Vocabulary and Reading Comprehension). Specifically, items from those sections that had been precalibrated were used as a basis for determining the estimated IRT curves. Then, chi-square statistics were computed to determine the degree to which the IRT curves, or "item ability regressions," matched the observed proportions of examinees (at each of several different ability levels) responding correctly to each item. (This is the standard procedure used to scale noncalibrated TOEFL items; cf. Hicks, 1984.)

For the analysis of the IRT data, three sets of parameters were estimated for the MC cloze items, based on a sample of 3,726 examinees:

- (a) MC cloze items scaled using TOEFL Section 2 (Structure and Written Expression) precalibrated items (CL2).
- (b) MC cloze items scaled using TOEFL Section 3 (Vocabulary and Reading Comprehension) precalibrated items (CL3)
- (c) MC cloze items scaled independently of any operational items (CLO).

Goodness of fit, graphical analysis. Following the standard procedures used with regular TOEFL items, item ability regressions of



the MC cloze items were visually compared with the curve resulting from the observed proportions of examinees at selected ability levels responding correctly to each item. The graphical data for the sets of parameter estimates based on Section 2 and Section 3 (CL2 and CL3) were evaluated. Typical examples of these item ability regressions are presented in Figures 1 and 2. (Both figures represent the same items, and item numbers represent location in the item calibration set.)

For both sections, the fit of the curve appeared to be excellent for all items except item 4, which was poorly fit in both sections (see Figures 3 and 4). The average chi-square value for CL2 was 19.92 and for CL3, 18.64 (see Table 11). These can be compared with average values of chi-square for the operational items in Sections 2 and 3, which were 20.24 and 20.96, respectively. The slightly higher values observed for the operational items reflect the fact that for some items (the scaling items), b parameters were being held fixed, resulting in slightly greater deviations for the theoretical curve. The results of this analysis indicated that the MC cloze items fit the IRT model at least as well as the operational items, and that the parameters based on Section 3 precalibrated items provided a closer fit to the model.

Bejar's test for unidimensionality. A test of the unidimensionality of a set of items as given by Bejar (1980) requires the comparison of b parameters derived from scaling the items in the context of a total test (i.e., CL2 and CL3) and b parameters derived by calibrating the items independently (CL0). Since both sets of parameters are based on the same students and the same items, they define an underlying metric; consequently, the item parameters should not differ unless the MC cloze items are assessing a unique ability not shared by either Section 2 or 3. If unidimensionality holds, the principal axis of the plot of the cloze item b parameter estimates should be close to unity with an intercept of zero. To assess this criterion, the slope and intercept of the principal axis are computed:

$$B = ((\sigma_1^2 - \sigma_2^2) + [(\sigma_1 - \sigma_2)^2 + 4\sigma_{12}^2]^{1/2}) / 2\sigma_{12}.$$

$$a = \bar{Y}_1 - B\bar{Y}_2.$$

where the subscripts, 1 and 2, index the statistics related to the total test difficulty estimates (CL2 and CL3) and the independent estimates (CL0), respectively. Unidimensionality is assessed in terms of the deviation of the angle of the principal axis from 45 degrees. The angle of the principal axis is derived from:

$$\tan 2\phi = \frac{2\sigma_{12}}{\sigma_1^2 - \sigma_2^2}.$$

Figure 1. Examples of item ability regressions for cloze items scaled with Section 2.

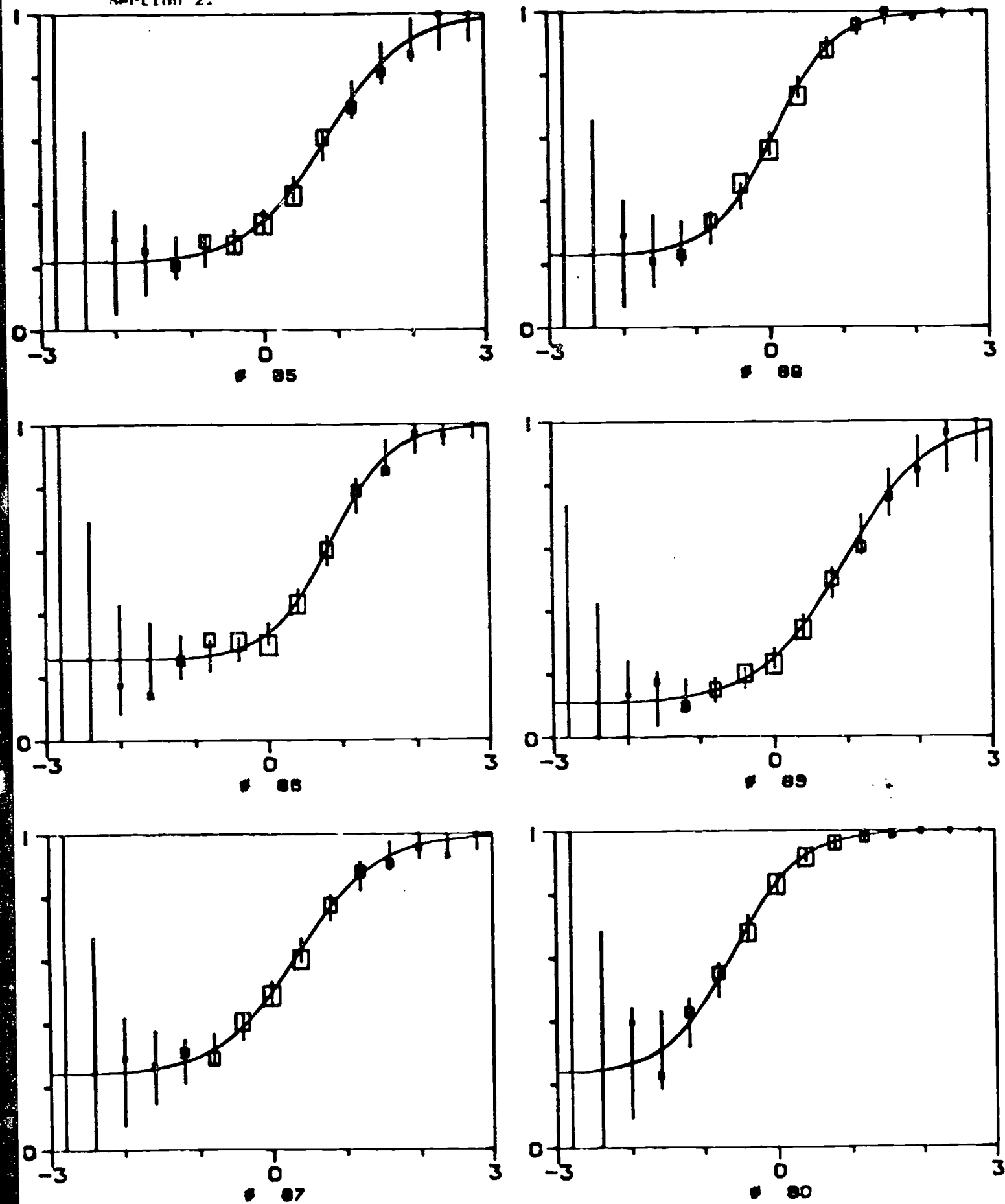


Figure 2. Examples of item ability regressions for cloze items scaled with Section 3.

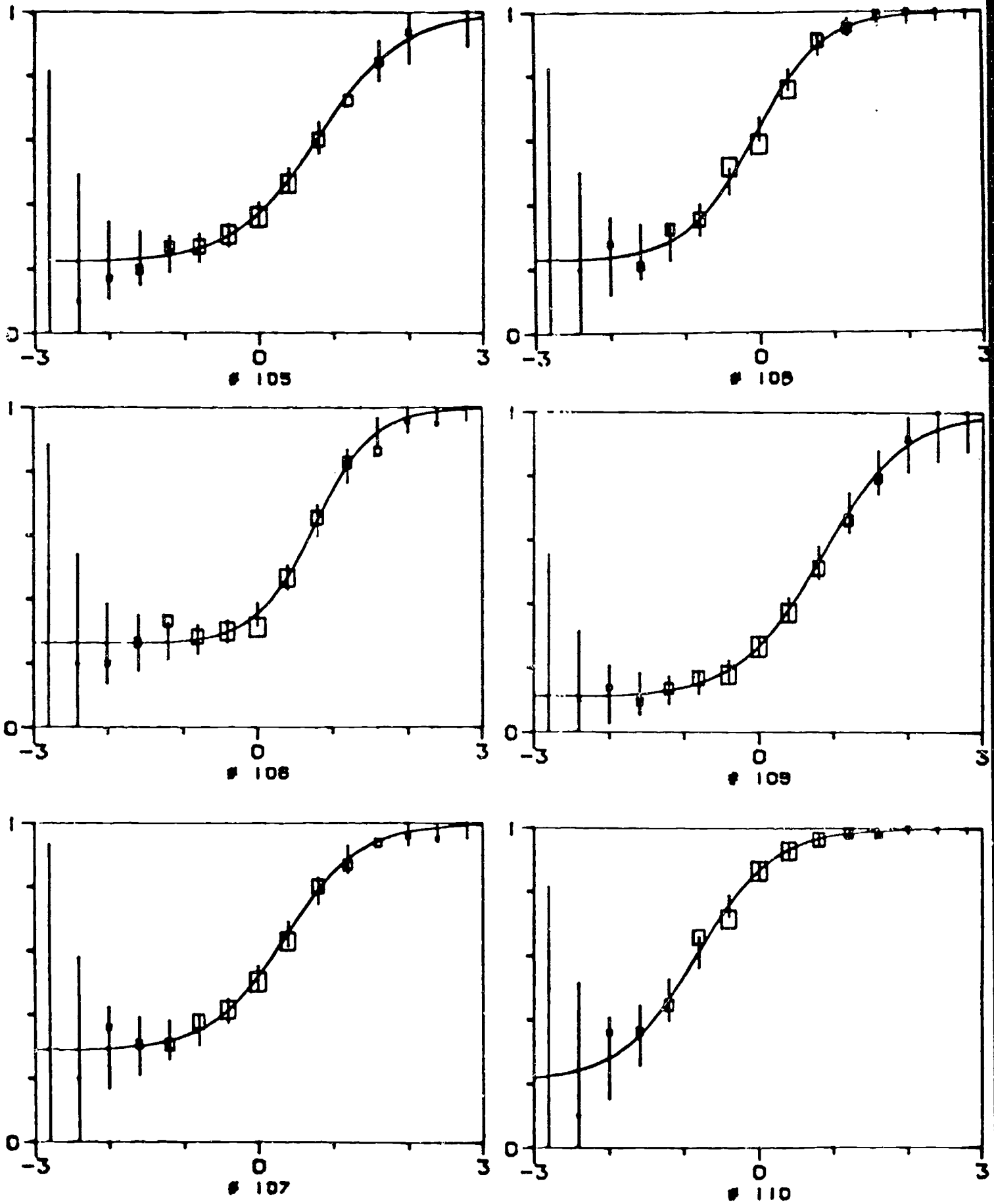


Figure 3. Item ability regressions for cloze item #4 scaled with Section 2.

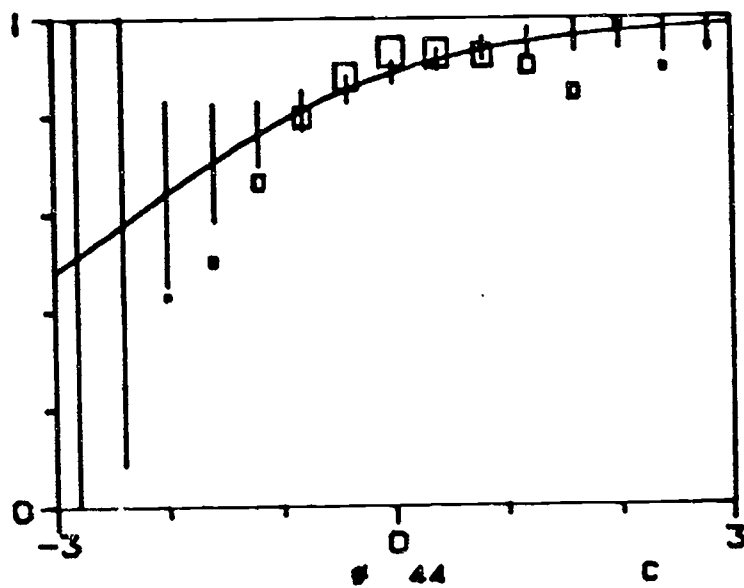


Figure 4. Item ability regression for cloze item #4 scaled with Section 3.

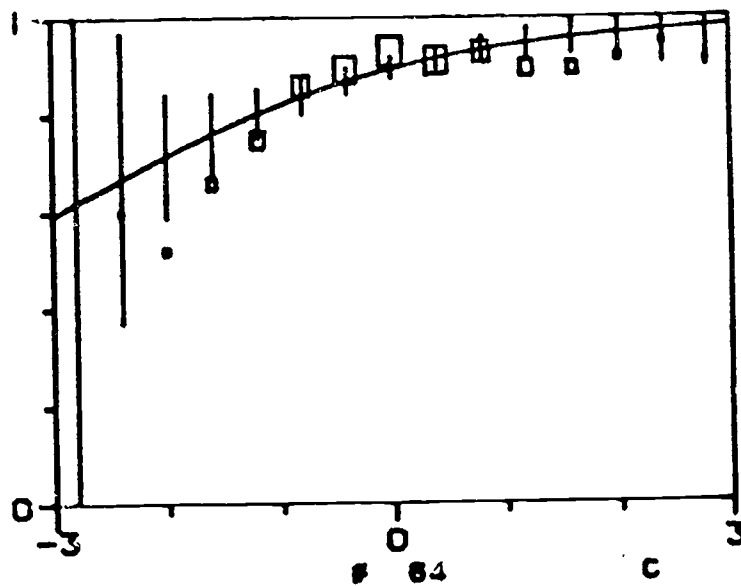


Table 11

Average Chi-Square Values for Cloze and  
Operational Items

	Section 2	Section 3
Cloze items	19.92	18.64
Operational items	20.24	20.96

The variance covariance matrices for the data in this study were:

Section 2			Section 3		
	1	2		1	2
1	$\begin{bmatrix} .71 & .75 \\ .75 & .83 \end{bmatrix}$		1	$\begin{bmatrix} .79 & .80 \\ .80 & .83 \end{bmatrix}$	
2			2		
	B = .93			B = .97	
	a = .18			a = .07	
	r = .97			r = .98	

Means and standard deviations of the b parameters were:

	Mean	S.D.
CL2	.11	.84
CL3	.00	.89
CLO	-.07	.91

The angle of the principal axis for the items scaled with Section 2 was 42.70 degrees, while for those scaled with Section 3, it was 44.23 degrees, thus indicating greater support for unidimensionality in terms of the latent trait assessed by Section 3. Plots of the b parameters with their principal axes are given in Figures 5 and 6.

Distributions and summary statistics of the a, b, and c parameters were also produced for each of the three MC cloze item calibrations (CL2, CL3, and CLO). These distributions were compared with operational data for Sections 2 and 3 (see Tables 12 and 13). The means of the a parameters for all three MC cloze item calibrations were less than those for the operational items in Sections 2 and 3, indicating that the MC cloze items were slightly less discriminating than the regular TOEFL items.

The mean b parameter for the MC cloze items scaled separately, or with Section 2 or 3, indicated that these items were more difficult than the items in either section (Table 13). The differences in the means obtained when the items were scaled with the two sections reflect the differences in the two scales. Relative to the Section 2 scale, the MC cloze items were quite difficult but seemed to be more reflective of the Section 3 scale.

Equating. An equating was performed for the 50 MC cloze items using the parameters obtained with the Section 3 precalibrated items (CL3). The MC cloze items were converted to the Section 3 scale using the standard TOEFL IRT equating program. The results produced a

Figure 5. Principal axis plot of CL0 and CL2.

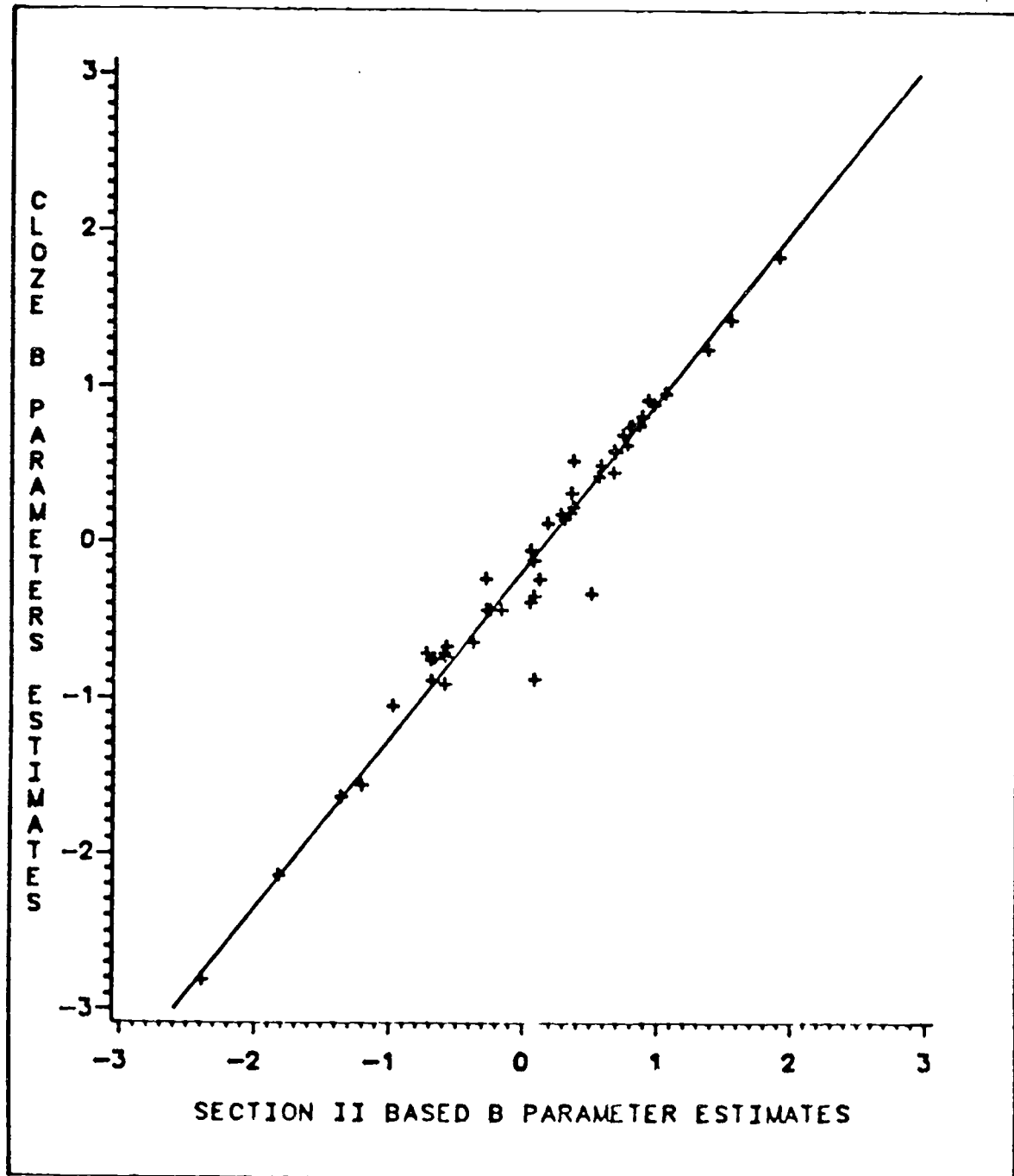


Figure 6. Principal axis plot of CL0 and CL3.

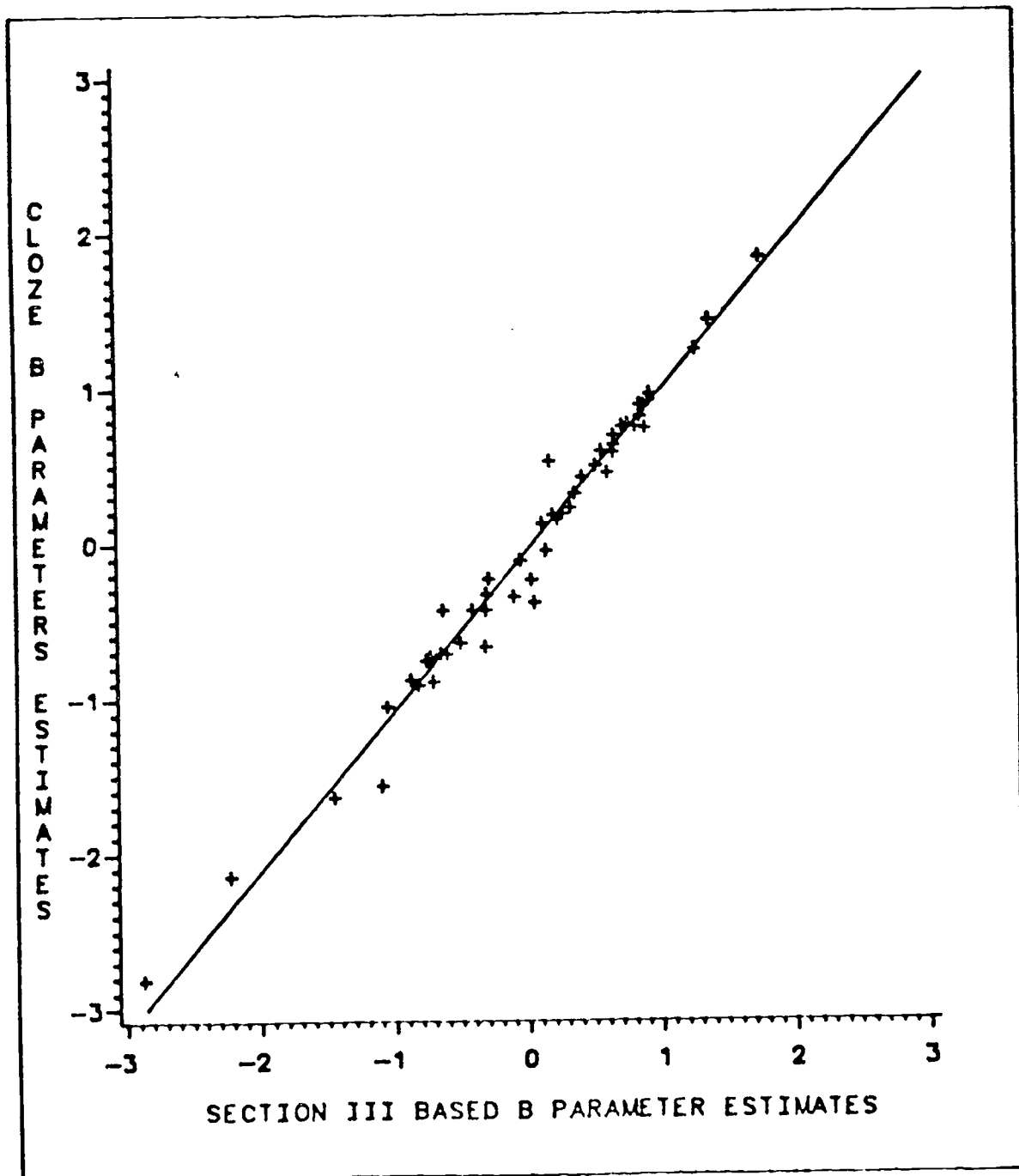




Table 12

## Frequency Distributions of a parameters

	<u>Section 2</u>	<u>Section 3</u>	<u>CL2</u>	<u>CL3</u>	<u>CL0</u>
1.41- 1.50	13	10	11	8	6
1.31- 1.40	4	5	2	2	2
1.21- 1.30	5	6	4	5	2
1.11- 1.20	1	3	2	5	8
1.01- 1.10	2	10	9	10	8
.91- 1.00	4	8	2	1	3
.81- .90	3	9	11	3	7
.71- .80	3	5	3	6	5
.61- .70	0	0	2	5	4
.51- .60	2	1	2	0	2
.41- .50	0	1	2	4	2
.31- .40	0			1	1
.21- .30	0				
.11- .20	1				
Total	38	58	50	50	50
Mean	1.16	1.09	1.05	1.01	1.02
S. D.	0.33	0.27	0.31	0.32	0.30

Table 13

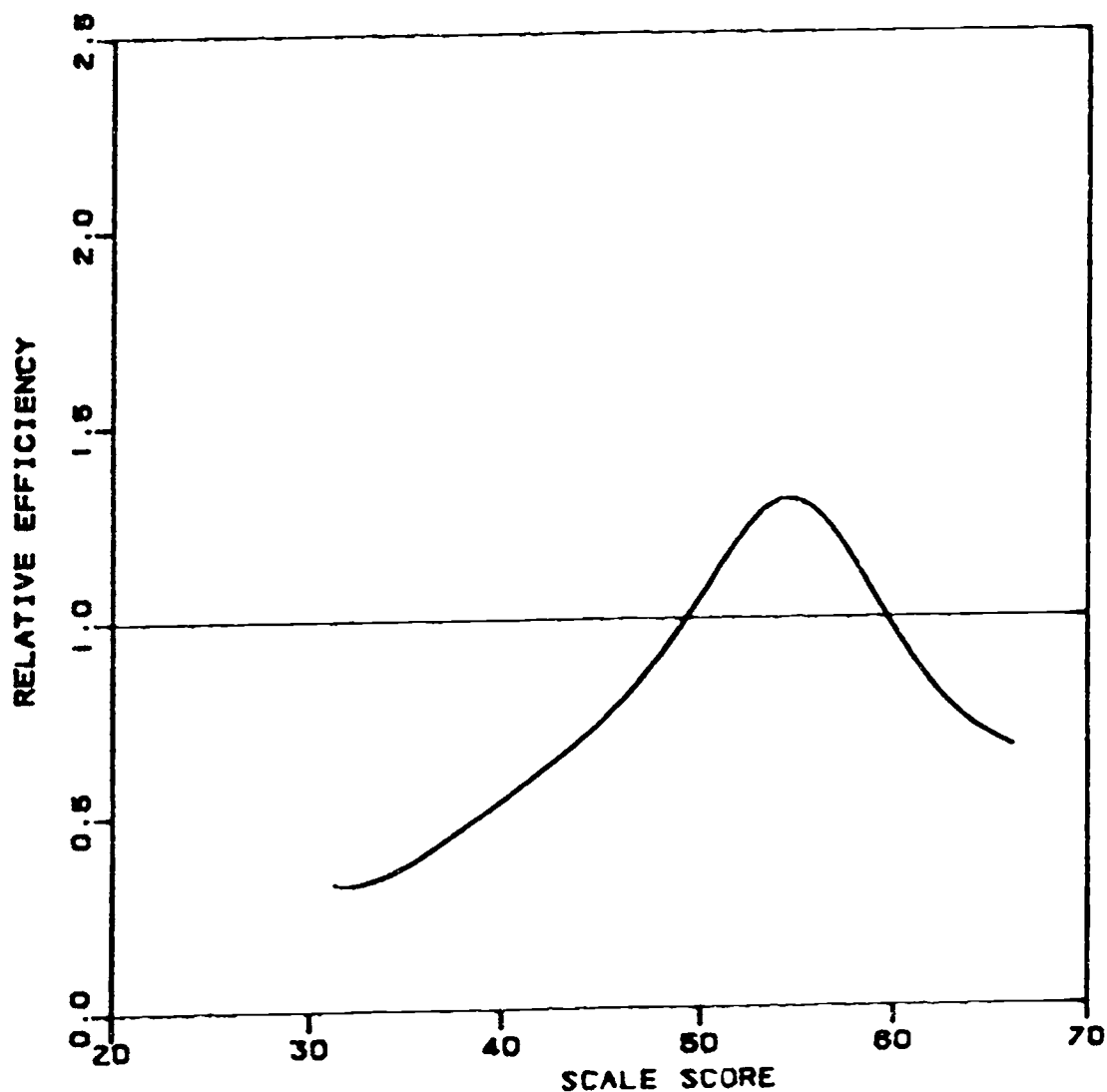
## Frequency Distributions of b parameters

		<u>Section 2</u>	<u>Section 3</u>	<u>CL2</u>	<u>CL3</u>	<u>CL0</u>
1.51-	2.00			2	1	1
1.01-	1.50	2	7	2	2	2
.51-	1.00	2	5	14	14	13
.01-	.50	9	16	14	11	9
-.49-	.00	9	17	5	7	9
-.99-	-.50	10	8	9	9	10
-1.49-	-1.00	2	1	2	3	1
-1.99-	-1.50	3	4	1	1	2
-2.49-	-2.00	0		1	1	2
-2.99-	-2.50	1				1
Total		38	58	50	50	50
Mean		-0.35	-0.12	0.11	-0.00	-0.07
S.D.		0.85	0.77	0.84	0.89	0.91

possible converted score range of 20 to 67 corresponding to a 0 to 50 raw score range. The equating curve was similar to those obtained operationally for the 58 items in Section 3. A plot of the relative efficiency of the MC cloze item test (relative to the TOEFL base form) was comparable to those obtained for regular forms of Section 3 (see Figures 7 and 8).

Interpretation. The items consisted mainly of two or three responses within a single sentence; that is, a single sentence could account for two or three items. Although lexical dependencies might exist among these responses, for statistical purposes, any such dependencies are apparently sufficiently weak so that IRT model assumptions regarding local independence are probably satisfied. This can be inferred from the tests of unidimensionality (graphical fit of the data to the model and Bejar's test of unidimensionality) that support the assumption of a unidimensional latent trait and, by implication, local independence. Evaluation of the graphical data provided support for unidimensionality with either Section 2 or Section 3. Bejar's test for unidimensionality indicated that a stronger relationship existed between the MC cloze items and Section 3 than with Section 2. Within the limits of these analyses, the operational use of the MC cloze item type in an IRT-based test such as the TOEFL would appear to be feasible.

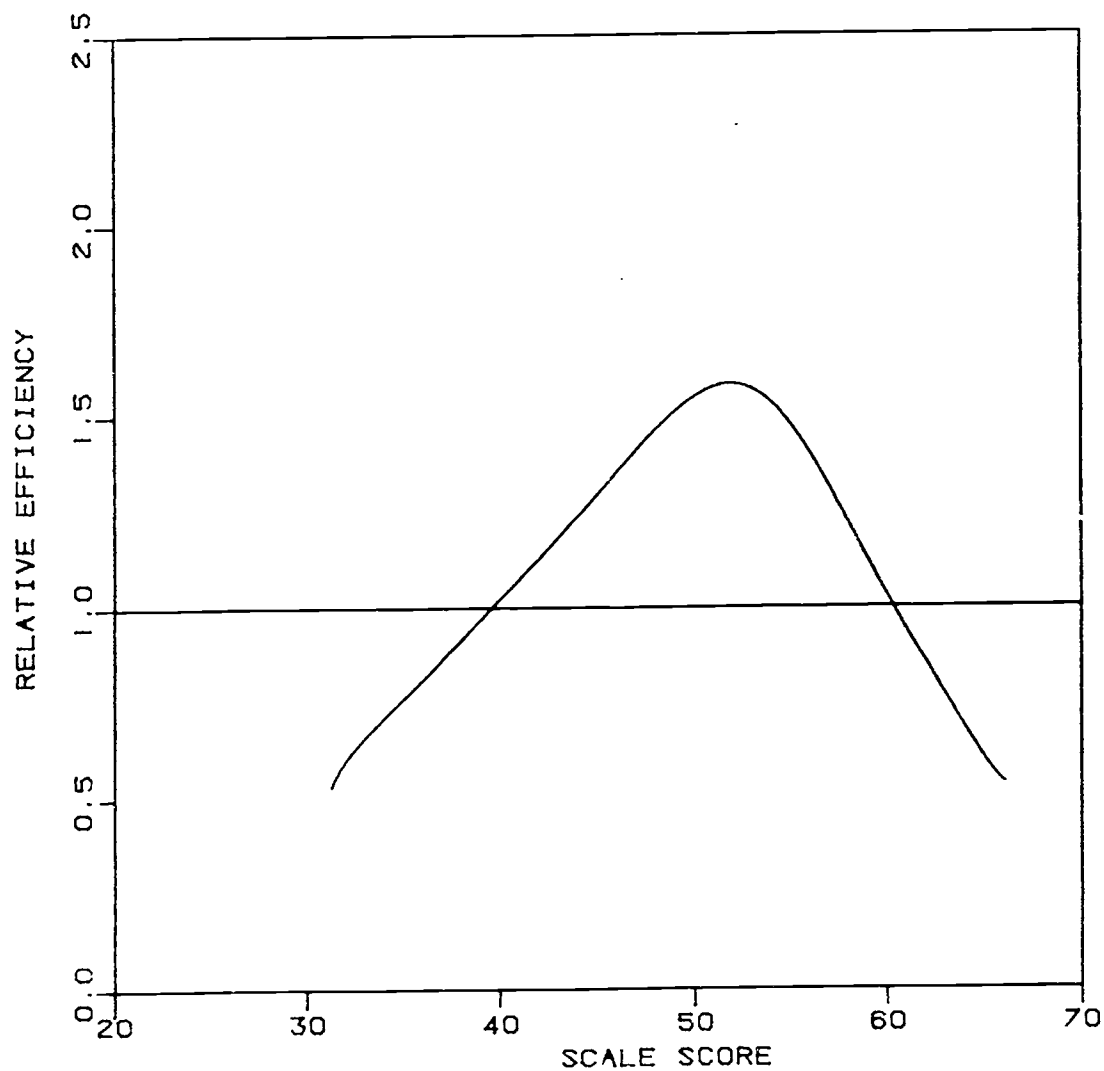
Figure 7. Relative efficiency for the test composed of cloze items.



READING COMPREHENSION AND VOCABULARY  
FORM CLOZE3 CZ SECTION 3  
ADM. DATE 3/85 PRE-EQUATING  
EFFICIENCY RELATIVE TO YTF3

Figure 8. Typical relative efficiency curve for TOEFL Section 3.

READING COMPREHENSION AND VOCABULARY  
FORM 3JTF5 JA SECTION 3  
ADM. DATE 5/87 N = 2971  
EFFICIENCY RELATIVE TO YTF3



## Discussion

The first issue investigated here concerned the factor structure of the basic TOEFL--that is, the principal three sections of the test. The analyses suggested that, for practical purposes, two basic factors are sufficient to account for performance on the test: one related to Listening Comprehension, and the other related to the remaining sections of the test (Structure, Written Expression, Vocabulary, and Reading Comprehension). This conclusion was suggested by the facts that, in confirmatory factor analyses, (a) goodness of fit indices were high for the two-factor solution, (b) root mean square residuals were low for the two-factor solution, and (c) correlations were high among the nonlistening factors in the three- and five-factor solutions, whereas correlations were lower between the Listening Comprehension factor and the other factors.

Actually, the benefit derived in going from one factor to two was not substantial, which suggests that the bulk of the variance in TOEFL performance is attributed to general proficiency (at least, in the receptive mode). At the same time, the evidence shows that the skills measured by the Listening Comprehension section, although related to other receptive language skills, are still somewhat separate from those measured by the other parts of the test.

The conclusion that fewer than three factors are sufficient to account for TOEFL performance, while consistent with the results of some studies (e.g., Hosley & Meredith, 1979; Manning, 1987; Oller & Hinofotis, 1980), is at odds with results obtained by Swinton and Powers (1980), which suggested a three-factor solution (albeit with slight differences in definition of the three factors for different language groups). A study is currently underway (Hale & Rock, in progress) to investigate the basis for the apparent discrepancy between the present findings and those of Swinton and Powers. Several hypotheses are being investigated. One hypothesis is that the present results are unique to domestic test takers (the population used in the present study), and that an examination of the entire population of both domestic and overseas TOEFL takers, as was done by Swinton and Powers, might produce different results. A second hypothesis is that differences in factor-analytic methodology play a role. Whereas the present study used item parcels as the units of analysis, Swinton and Powers used individual items, and it has been noted that use of individual items can sometimes produce a potentially spurious factor, such as one associated with item difficulty (Hulin, Drasgow, & Parsons, 1983). A third hypothesis is that the test had changed between 1976 and 1984 in the direction of greater overlap in the skills measured by the nonlistening subtests. Until the results of the follow-up study are available, it would be premature to speculate on the possible basis for the discrepancy in findings of these studies.

It should be noted that Oltman and Stricker (1988) have completed

a cluster analysis of the TOEFL and found that, if one examines low-difficulty items only, the items appear to fall into three distinct clusters that relate to the three sections of the test; however, if one examines medium- or high-difficulty items, the items fall into a single cluster. These results, along with the results of the study currently underway, are expected to provide a more comprehensive picture of the test's underlying structure.

Turning to the MC cloze items, an initial issue was whether it is possible to write MC cloze items that represent the different nonaural areas of English proficiency--vocabulary, reading, and grammar. It was determined that these areas are not entirely separable and concluded that, as a tentative model at least, MC cloze items are best characterized by a classification scheme involving combinations of these three areas, as defined above.

To empirically assess the distinctions among the categories in this scheme, correlations were computed among scores for the four categories of MC cloze items. The correlations proved to be about the same in magnitude as the reliabilities (suggesting that, if the correlations were corrected for unreliability, they would be near unity). Thus, the correlational data provide little empirical basis for concluding that the four categories of MC cloze items measure distinct aspects of language proficiency. It is possible that the kind of text processing required by cloze items in general may be more holistic in character than current linguistic analyses might suggest. In order for the reader to deal with a question involving grammar, for example, vocabulary must play some role; and one cannot answer a question involving vocabulary without a certain control of grammar. Furthermore, operating within clause boundaries may involve some of the same cognitive machinery and intuitions as operating across clause boundaries. So it is perhaps not entirely surprising that there is a commonality across the item types examined here.

A principal issue under study concerned the role of the MC cloze items in the TOEFL. More specifically, the objective was to determine how each category of MC cloze item related to each of the different parts of the TOEFL.

The correlations with MC cloze scores were generally lower for Listening Comprehension than for the other four parts of the TOEFL. This is consistent with the factor analytic results and reflects the distinction in skills measured by the Listening Comprehension section and the remainder of the test.

Regarding the nonlistening parts of the TOEFL, there was a slight tendency, relatively consistent across languages, for the RG and GR cloze items to correlate more highly with TOEFL Structure or Written Expression, and for the RV and VR cloze items to correlate more highly with TOEFL Vocabulary or Reading Comprehension. However, the differences observed were small. Thus, from a practical standpoint, the data do not support the notion that the different MC cloze items

are linked to distinct skills, as measured by the TOEFL.

It is not surprising that the relations between MC cloze item categories and parts of the TOEFL did not differ substantially. Factor analysis of the TOEFL did not provide evidence that the Structure, Written Expression, Vocabulary, and Reading Comprehension parts measure separate skills. Also, correlations among scores for the MC cloze item types were relatively high compared with their reliabilities, providing little evidence of a difference in skills tapped by the four cloze item categories. Thus, given that neither the internal analysis of the TOEFL nor the internal analysis of the cloze test indicated measurement of distinct nonaural skills, differential relations between the various MC cloze categories and the various nonaural parts of the TOEFL would not be expected.

In multiple regression analyses with the total MC cloze score as the dependent variable and the five TOEFL part scores as the independent factors (conducted separately per language group), very high multiple correlations were obtained--mostly in the low to high .90s. This analysis suggests that total MC cloze performance can be almost completely predicted by scores on the TOEFL, so that the present MC cloze items may add little to measurement of English proficiency beyond that already provided by the TOEFL. (Note, however, that the reverse is not necessarily the case; indeed, listening comprehension appears to be an aspect of English proficiency tapped by the TOEFL that is not measured by the MC cloze items.)

Finally, analyses were performed to determine whether the IRT parameters could be adequately estimated from the MC cloze data. The three parameters, in effect, reflect an item's discrimination power, difficulty, and guessing level. The analyses, performed for all language groups combined, indicated that these parameters could indeed be satisfactorily estimated, with quite good fit. To test the assumption of local independence of the MC cloze items would be an extremely difficult matter and was beyond the scope of the present study. However, the excellent fit of the three IRT parameters observed here, as well as the results of Bejar's test of unidimensionality, indirectly imply that items are sufficiently independent for all practical purposes and suggest that IRT methods can appropriately be applied to the MC cloze items, just as they are applied to other item types used in the TOEFL.

### Practical Implications

The data obtained in the present study suggested that, from a practical standpoint, TOEFL performance seems to be characterized by two factors, defined as (a) Listening Comprehension and (b) the combination of Structure, Written Expression, Vocabulary, and Reading Comprehension. Also, the MC cloze item types were relatively highly related to each other. Partly because of these last two facts, there was no strong indication of differential relations between the four MC



cloze item types and the four nonlistening parts of the TOEFL. Thus, contrary to the hypothesized relationships discussed at the outset of this report, the study did not identify MC cloze items that were shown to assess reading comprehension apart from other skills in the receptive, nonaural domain. Furthermore, there appeared to be a high degree of overlap in the skills measured by the MC cloze items and those already measured by the nonlistening parts of the TOEFL.

Despite these results, there may still be reasons for considering the MC cloze format as a component of an English proficiency test such as the TOEFL. The cloze methodology is regarded by many as providing an "integrative" mode of assessment, which essentially taps aspects of English proficiency in a broader context than do "discrete-point" items. Given that some TOEFL items--such as those in the Vocabulary subsection--are not as integrative as MC cloze items, it might be argued that inclusion of the latter items on the TOEFL could help broaden the scope of measurement techniques employed on the test.<sup>6</sup> Furthermore, as seen in this study, the MC cloze procedure seems to have a reliability that is comparable to that of the TOEFL, adjusting for differences in number of items. And MC cloze items appear to be amenable to IRT parameter estimation, thus making them eligible for inclusion in a test that relies on IRT equating methods, such as the TOEFL.

Apparently, several factors need to be weighed in considering the possible use of MC cloze items for measuring reading comprehension on the TOEFL. In particular, theoretical considerations such as those just discussed, as well as the empirical evidence, should be taken into account. Regarding the latter, while there were some negative practical implications of the present results, there are certain issues that remain unresolved and could be addressed in further research. For example, although the present MC cloze classification scheme did not produce a subset of items that clearly involved reading comprehension more than other aspects of English proficiency, exploration of alternative classification schemes could yield different results. Also, although the present study focused on the internal structure of the TOEFL and the role of MC cloze items therein, it would also be of value to study the relation of MC cloze items and TOEFL items to external criterion measures. Further, it needs to be determined how easily MC cloze items could be developed on a routine basis by test development specialists. Investigation of issues such as these could provide a basis for more informed decision making about the role of MC cloze items in large-scale assessment of English proficiency.

---

<sup>6</sup>The multiple-choice version of the cloze procedure, although lacking the constructive-response feature of the completion cloze format, is still regarded as integrative in the sense that it requires the examinee to deal with several linguistic components at the same time.

## References

- Alderman, D. L., & Holland, P. W. (1981). Item performance across native language groups on the Test of English as a Foreign language (TOEFL Research Rep. No. 9; ETS Research Rep. No. 81-16). Princeton, NJ: Educational Testing Service.
- Alderson, J. C. (1983). The cloze procedure and proficiency in English as a foreign language. In J. W. Oller, Jr. (Ed.), Issues in language testing research. Rowley, MA: Newbury House.
- Anderson, J. (1976). Psycholinguistic experiments in foreign language testing. St. Lucia, Queensland, Australia: University of Queensland Press.
- Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. TESOL Quarterly, 19, 535-556.
- Baldauf, R. B., Jr., & Propst, I. K., Jr. (1979). Matching and multiple-choice cloze tests. Journal of Educational Research, 72, 321-326.
- Bejar, I. I. (1980). A procedure for investigating the unidimensionality of achievement tests based on item parameter estimates. Journal of Educational Measurement, 17, 283-296.
- Bensoussan, M. (1983). Multiple-choice modifications of the cloze procedure using word-length and sentence-length blanks. Foreign Language Annals, 16, 189-198.
- Bensoussan, M., & Ramraz, R. (1984). Testing EFL reading comprehension using a multiple-choice rational cloze. Modern Language Journal, 68, 230-239.
- Bormuth, J. R. (1966, February). Factor validity of cloze tests as measures of reading comprehension ability. Paper presented at meeting of the American Educational Research Association.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. Modern Language Journal, 64, 311-317.
- Carroll, J. B. (1961). Fundamental considerations in testing for English proficiency of foreign students. In Center for Applied Linguistics, Testing the English proficiency of foreign students. Washington, DC: Author.
- Cranney, A. G. (1973). The construction of two types of cloze reading tests for college students. Journal of Reading Behavior, 5, 60-64.
- Educational Testing Service (1987). TOEFL test and score manual. Princeton, NJ: Author.

- Friedman, M. M. (1964). The use of the cloze procedure for improving the reading comprehension of foreign students at the University of Florida. Unpublished doctoral dissertation, University of Florida.
- Guthrie, J. T. (1973). Reading comprehension and syntactic responses in good and poor readers. Journal of Educational Psychology, 65, 294-299.
- Hale, G. A., & Rock, D. A. Factor structure of the TOEFL: A further investigation. (Project in progress.)
- Hicks, M. M. (1984). A comparative study of methods of equating TOEFL test scores. (ETS Research Rep. No. 84-20). Princeton, NJ: Educational Testing Service.
- Hinofotis, F. B. (1978). Cloze as an alternative method of ESL placement and proficiency testing. In J. W. Oller, Jr., & K. Perkins (Eds.), Research in language testing. Rowley, MA: Newbury House.
- Hinofotis, F. B., & Snow, B. G. (1978). An alternative cloze testing procedure: Multiple-choice format. In J. W. Oller, Jr., & K. Perkins (Eds.), Research in language testing. Rowley, MA: Newbury House.
- Hosley, D., & Meredith, K. (1979). Inter- and intra-test correlates of the TOEFL. TESOL Quarterly, 13, 209-217.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). Item response theory. Homewood, IL: Dow Jones-Irwin.
- Irvine, P., Atai, P., & Oller, J.W., Jr. (1974). Cloze, dictation, and the Test of English as a Foreign Language. Language Learning, 24, 245-252.
- Jonz, J. (1976). Improving on the basic egg: The M-C cloze. Language Learning, 26, 255-265.
- Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.
- Joreskog, K. G., & Sorbom, D. (1981). LISREL V--Analysis of linear structural relationships by maximum likelihood and least squares methods. Chicago: International Educational Services.
- Joreskog, K. G., & Sorbom, D. (1983). LISREL VI--Supplement to the LISREL V manual. Uppsala, Sweden: University of Uppsala, Department of Statistics.
- Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

- Manning, W. H. (1987). Development of cloze-elide tests of English as a second language. (TOEFL Research Rep. No. 23; ETS Research Rep. No. 87-18). Princeton, NJ: Educational Testing Service.
- Oller, J. W., Jr. (1979). Language tests at school: A pragmatic approach. London: Longman.
- Oller, J. W., Jr., & Hinofotis, F. B. (1980). Two mutually exclusive hypotheses about second language ability: Indivisible or partially divisible competence. In J. W. Oller, Jr., & K. Perkins (Eds.), Research in language testing. Rowley, MA: Newbury House.
- Oltman, P. K., & Stricker, L. J. (1988). Native language, English proficiency, and the structure of the Test of English as a Foreign Language. (TOEFL Research Rep. No. 27). Princeton, NJ: Educational Testing Service.
- O'Reilly, R. P., & Schuder, R. T. (1977, April). Some issues in the measurement of basic competence in reading. Paper presented at the meeting of the American Educational Research Association, New York.
- O'Reilly, R. P., & Streeter, R. E. (1977). Report on the development and validation of a system for measuring literal comprehension in a multiple-choice cloze format: Preliminary factor analytic results. Journal of Reading Behavior, 9, 45-69.
- Pike, L. W. (1979). An evaluation of alternative item formats for testing English as a foreign language (TOEFL Research Rep. No. 2; ETS Research Rep. No. 79-6). Princeton, NJ: Educational Testing Service.
- Scholz, G. E., & Scholz, C. M. (1981, Detroit). Multiple-choice cloze tests of EST discourse: An exploration. Paper presented at the 15th annual TESOL convention.
- Swinton, S. S., & Powers, D. E. (1980). Factor analysis of the Test of English as a Foreign Language for several language groups. (TOEFL Research Rep. No. 6; ETS Research Rep. No. 80-32). Princeton, NJ: Educational Testing Service.
- Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. Journalism Quarterly, 30, 415-433.

## Appendix A

TOEFL Test Form Used in Study, Including Fourth, Cloze Section

## SECTION 1

## LISTENING COMPREHENSION

In this section of the test, you will have an opportunity to demonstrate your ability to understand spoken English. There are three parts to this section, with special directions for each part.

## Part A

Directions: For each question in Part A, you will hear a short statement. The statements will be spoken just one time. They will not be written out for you, and you must listen carefully to understand what the speaker says.

After you hear a statement, read the four sentences in your test book, marked (A), (B), (C), and (D), and decide which one is closest in meaning to the statement you heard. Then, on your answer sheet, find the number of the question and blacken the space that corresponds to the letter of the answer you have chosen so that the letter inside the oval cannot be seen.

## Example 1

You will hear:

You will read:

- (A) John does better in his studies than James.  
(B) James is bigger than his brother John.  
(C) John has only one brother.  
(D) The teacher likes James better than John.

Sentence (A), "John does better in his studies than James," means most nearly the same as the statement "John is a better student than his brother James." Therefore, you should choose answer (A).

## Example 11

You will hear:

You will read:

- (A) The traffic isn't bad today.  
(B) The trucks weigh a lot.  
(C) There are a lot of trucks on the highway.  
(D) The highway has been closed to heavy trucks.

Sentence (C), "There are a lot of trucks on the highway," is closest in meaning to the sentence "The truck traffic on this highway is so heavy I can barely see where I'm going." Therefore, you should choose answer (C).

- 1 (A) The dinner wasn't for Susan or Fred.  
(B) Neither Fred nor Susan came to dinner.  
(C) Susan wanted to meet Fred for dinner.  
(D) Fred asked Susan to help him make dinner.

- 2 (A) We will attend the program.  
(B) It will go beyond the tent.  
(C) There will be ten grams of it.  
(D) It will begin at ten o'clock.

3. (A) Keith is one of Ellen's employees.  
(B) Keith is very dedicated.  
(C) Ellen is the better employee.  
(D) Ellen is dictating a letter to Keith.

4. (A) I think you should have an open mind.  
(B) Why aren't the windows open?  
(C) Please leave the window closed.  
(D) Didn't you mind doing the work?

5. (A) The key is gone.  
(B) The secretary has locked the door.  
(C) The key is in the laboratory.  
(D) The laboratory equipment has been stolen.

6. (A) Exam questions are never easy.  
(B) After reviewing the exam, I found nothing easy in it.  
(C) Once your exam is returned to you, the questions seem easy.  
(D) The questions on the back of the exam were easier than the others.

7. (A) I'm gradually improving my French.  
(B) I bit my tongue while speaking French.  
(C) I'm not taking enough French.  
(D) I've been writing my notes in French.

8. (A) The shoe store is on the right.  
(B) Try on the right shoe, please.  
(C) I can't find the shoe store.  
(D) Tell me if that's the shoe store.

9. (A) Rose is a hair stylist.  
(B) Rose was reading a newspaper.  
(C) Rose was editing a book.  
(D) Rose is a teacher.

10. (A) It's crowded today.  
(B) It was far away.  
(C) He has many doubts.  
(D) The sky was overcast.

11. (A) Robert wasn't aware of the evidence.  
(B) The evidence didn't injure Robert.  
(C) Robert wouldn't consider the evidence.  
(D) The evidence was incomplete.

12. (A) What is the carpet made of?  
(B) What was it you spilled?  
(C) You spoiled what was there.  
(D) You spilled it in the car.

13. (A) I'm ready to see that couple.  
(B) I'm already married.  
(C) I have two references now.  
(D) I have referees who are prepared.

14. (A) We would like to go to the ballet more frequently than we do now.  
(B) Our ballet company has recently been appearing at that theater.  
(C) Lately, we've been attending theater performances rather than the ballet.

- (D) We don't go to the theater as frequently as we used to.

15. (A) They didn't suspect that I already knew the answer.  
(B) I suspected that they knew very few answers.  
(C) They suspected me because I knew the answer.  
(D) They suspected my answer from the start.

GO ON TO THE NEXT PAGE

170

GO ON TO THE NEXT PAGE

170

16. (A) Sam probably had a skiing accident.  
(B) Sam tripped and fell on Tuesday.  
(C) Sam is on the last leg of his trip.  
(D) Sam is probably skiing today.
17. (A) They showed us the clothes.  
(B) The shore is close.  
(C) The market was shut down.  
(D) The exhibit has ended.
18. (A) How tall are you?  
(B) You have become very tall.  
(C) You have to grow some more.  
(D) How did you measure the height?
19. (A) We're free to study drama and poetry, too.  
(B) We were told to study the novel.  
(C) We can take some of the branches off.  
(D) We want to live in another area.
20. (A) I can't think when my head aches.  
(B) I didn't get any headaches when I went away.  
(C) My headache doesn't seem to be getting any better.  
(D) It took a while, but my head finally stopped hurting.

### Part B

Directions: In Part B you will hear short conversations between two speakers. At the end of each conversation, a third voice will ask a question about what was said. The question will be spoken just one time. After you hear a conversation and the question about it, read the four possible answers in your test book and decide which one is the best answer to the question you heard. Then, on your answer sheet, find the number of the question and blacken the space that corresponds to the letter of the answer you have chosen.

#### Example 1

You will hear:

You will read:

- (A) Read a book.  
(B) Write a composition.  
(C) Talk about a problem.  
(D) Listen to the radio.

#### Sample Answer

Ⓐ Ⓑ Ⓒ Ⓓ

From the conversation you know that the assignment is to listen to a radio program and be ready to talk about it. The best answer, then, is (D), "Listen to the radio." Therefore, you should choose answer (D).

21. (A) The train will be heavily loaded.  
(B) The Capitol Building is made of stone.  
(C) The Capitol Building is near the train station.  
(D) The train has already departed for Washington.
22. (A) He doesn't think they are allowed to speak.  
(B) He doesn't know what's happening outside.  
(C) He was only talking to himself.  
(D) He thinks it's too noisy to talk now.

23. (A) He can read for a long time when he's interested.  
(B) He's also amazed at how much he reads.  
(C) He reads the same amount as the woman does.  
(D) He finds it difficult to sit still to read.
24. (A) Buy a new desk.  
(B) Stand beside the desk.  
(C) Take up painting.  
(D) Move the picture.
25. (A) He's unhappy because he has to go to class.  
(B) He's making steady progress.  
(C) He's a good swimmer already.  
(D) He's sure the class is too slow for him.
26. (A) What she should wear.  
(B) Where the man lives.  
(C) Who will speak at the meeting.  
(D) Why a meeting is necessary.
27. (A) Take the refrigerator away.  
(B) Buy some groceries.  
(C) Pick out other appliances.  
(D) Eat lunch at work.
28. (A) He may have made a mistake.  
(B) He was looking for Harry in the library.  
(C) He's not sure which way to go.  
(D) He may see Harry in California.
29. (A) He got sick flying.  
(B) He just got a new flute.  
(C) He doesn't eat enough fruit.  
(D) He is no longer ill.
30. (A) She doesn't need a new roll of film yet.  
(B) She still hasn't finished the first film.  
(C) She'll soon have two rolls to be developed.  
(D) She'd like him to develop the film for her now.
31. (A) It's a very good bargain.  
(B) It's too expensive.  
(C) If you buy three, the price is lower.  
(D) You would pay the same price for it in other stores.
32. (A) Eliminate certain required classes.  
(B) Offer more basic classes.  
(C) Expel some students.  
(D) Require more student discussions.
33. (A) Visiting a zoo.  
(B) Making orange juice.  
(C) Looking for a room.  
(D) Packing a suitcase.
34. (A) He'll make a bet with President Blake.  
(B) President Blake would probably serve as master of ceremonies.  
(C) President Blake will be honored to receive the award.  
(D) He'll get President Blake's approval for the banquet.
35. (A) He didn't score very many points.  
(B) He didn't mind watching the basketball game.  
(C) He didn't know where the game was going to be held.  
(D) He didn't want to play basketball originally.



## Part C

Directions: In this part of the test, you will hear several short talks and conversations. After each talk or conversation, you will be asked some questions. The talks and questions will be spoken just one time. They will not be written out for you, so you will have to listen carefully to understand what the speaker says.

After you hear a question, read the four possible answers in your test book and decide which one is the best answer to the question you heard. Then, on your answer sheet, find the number of the question and blacken the space that corresponds to the letter of the answer you have chosen.

Listen to this sample talk.

You will hear:

Now look at the following example.

You will hear:

- You will read: (A) By plane.  
(B) By ship.  
(C) By train.  
(D) By bus.

Sample Answer

Ⓐ ● Ⓑ ○

The best answer to the question "How did people generally arrive at Ellis Island?" is (B), "By ship." Therefore, you should choose answer (B).

Now look at the next example.

You will hear:

- You will read: (A) New immigrants.  
(B) International traders.  
(C) Fishermen.  
(D) Tourists.

Sample Answer

Ⓐ ○ Ⓑ ○ ●

The best answer to the question "Who visits Ellis Island today?" is (D), "Tourists." Therefore, you should choose answer (D).

36. (A) An artist.  
(B) A professor of library science.  
(C) A doctor specializing in vision.  
(D) A professor of art history.

39. (A) Those who have visual disabilities.  
(B) Those who intend to become artists.  
(C) Those who have no interest in painting.  
(D) Those who have never taken art history before.

37. (A) The history of the slide library.  
(B) The use of slides in the course.  
(C) The material to be tested that day  
(D) The outline of the course.

40. (A) Students.  
(B) Musicians.  
(C) Professors.  
(D) Accountants.

38. (A) The beginning.  
(B) Just before the first weekly exam.  
(C) Halfway through the course.  
(D) Just before the final exam.

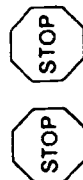


41. (A) They all play trumpets.  
(B) They all play modern jazz.  
(C) They are all over seventy.  
(D) They are all Southerners.
42. (A) She hasn't had dinner yet.  
(B) She needs a break from her studies.  
(C) She couldn't go to the last concert.  
(D) She knows the band is good.
43. (A) Lend the man some money.  
(B) Go to the bank for cash.  
(C) Pay for both tickets later.  
(D) Share the costs with the man.
44. (A) Appreciative.  
(B) Unhappy.  
(C) Excited.  
(D) Embarrassed.
45. (A) At the concert hall.  
(B) At the ticket booth.  
(C) At the dormitory.  
(D) At the restaurant.
46. (A) In A.D. 600.  
(B) During the third century.  
(C) More than three hundred years ago.  
(D) In the 1940's.
47. (A) Powdered cinnamon.  
(B) Tree bark.  
(C) Poplar leaves.  
(D) Crushed ore.
48. (A) Curing leather.  
(B) Treating deep cuts.  
(C) Curing a disease.  
(D) Preserving food supplies.
49. (A) Quinine dropped in value.  
(B) Cinchona trees almost disappeared.  
(C) Quinine was often mixed with other substances.  
(D) Cinchona supplies were cut off by the Peruvians.
50. (A) To make room for settlements.  
(B) To provide a fuel supply.  
(C) To make a road through Peru.  
(D) To get all the bark off.

THIS IS THE END OF THE LISTENING COMPREHENSION SECTION OF THE TEST

THE NEXT PART OF THE TEST IS SECTION 2. TURN TO THE DIRECTIONS FOR SECTION 2 IN YOUR TEST BOOK, READ THEM, AND BEGIN WORK.

DO NOT READ OR WORK ON ANY OTHER SECTION OF THE TEST.





## SECTION 2

## STRUCTURE AND WRITTEN EXPRESSION

Time - 25 minutes

This section is designed to measure your ability to recognize language that is appropriate for standard written English. There are two types of questions in this section, with special directions for each type.

**Directions:** Questions 1-15 are incomplete sentences. Four words or phrases, marked (A), (B), (C), and (D), are given beneath each sentence. You are to choose the one word or phrase that best completes the sentence. Then, on your answer sheet, find the number of the question and blacken the space that corresponds to the letter of the answer you have chosen so that the letter inside the oval cannot be seen.

Example 1

Mt. Hood ----- in the state of Oregon.

- (A) although  
(B) and  
(C) is  
(D) which

Sample Answer

☐ (A) ☒ (B) ☐ (C) ☐ (D)

In English, the sentence should read, "Mt. Hood is in the state of Oregon." Therefore, you should choose (C).

Example 2

----- most important event in San Francisco's history was the disastrous earthquake and fire of 1906.

- (A) The  
(B) It was the  
(C) That the  
(D) There was a

Sample Answer

☒ (A) ☐ (B) ☐ (C) ☐ (D)

In English, the sentence should read, "The most important event in San Francisco's history was the disastrous earthquake and fire of 1906." Therefore, you should choose (A).

As soon as you understand the directions, begin work on the questions.

1. The activities of the Tennessee Valley Authority have aided the economic rehabilitation of the Tennessee Valley, ----- of some 40,000 square miles.

- (A) an area  
(B) its area  
(C) area  
(D) areas

2. The pigment in a paint not only creates a decorative and functional color ----- affects the mechanical properties of the paint.

- (A) and then  
(B) that  
(C) but also  
(D) or

3. Thurgood Marshall ---- - practicing law in 1933.

- (A) began  
(B) was begun  
(C) beginning  
(D) he began

4. ----- were wood engravings.

- (A) Printing the units engraved earliest  
(B) Printing the earliest engraved units  
(C) The earliest engraved printing units  
(D) Engraved the earliest printing units

5. Thomas Paine was committed to speaking the truth bluntly, -----.

- (A) was understood  
(B) he had understanding  
(C) as he understood it  
(D) that his understanding was

6. -----, angiosperms have specialized conducting tissues.

- (A) Other vascular plants are like them  
(B) They are like other vascular plants  
(C) Other, like vascular plants  
(D) Like other vascular plants

7. The artist Scipio Moorhead ----- primarily in the poetry of Phillis Wheatley.

- (A) is remembered  
(B) being remembered  
(C) that it is remembered  
(D) remembered

8. Liberals tend to favor ----- do conservatives.

- (A) more than immediate social change  
(B) change more immediate than social  
(C) more immediate social change than  
(D) social change more than immediate

9. ----- are fed into a tape recorder, they magnetize the particles on the tape in varying patterns.

- (A) Electric waves  
(B) When electric waves  
(C) Because of electric waves  
(D) Electric waves that

10. Inaugurated a second time on March 4, 1901, ----- focused on domestic rather than foreign policies.

- (A) William McKinley's new term looked forward to and  
(B) the new term looked forward to William McKinley and  
(C) looking forward to a new term was William McKinley  
(D) William McKinley looked forward to a new term

11. Porcupines ----- a dense layer of wool next to their skin.

- (A) grow  
(B) growing  
(C) which grow  
(D) were being grown

12. Playwright Lillian Hellman's dramas are marked by their intelligent and -----.

- (A) weaved tight plots  
(B) plots are tightly woven  
(C) tightly woven plots  
(D) weaving of tight plots

13. Native to the western United States, mariposa lilies have narrow ----- like large blades of grass.

- (A) shape of leaves  
(B) leaves shape them  
(C) leaf-shaped  
(D) leaves shaped

GO ON TO THE NEXT PAGE

GO ON TO THE NEXT PAGE

- (A) What mathematics  
(B) Where mathematics  
(C) Mathematics  
(D) That mathematics

-10-

- 24 The marimba is a percussion instrument like the xylophone.

**GO ON TO THE NEXT PAGE**

25. Civilization today had depended A on wood B at great cost to the world's natural resources. D
26. No bank keeps enough cash B paying all its depositors in full at one time. A
27. The style of a play derives from the A specify material and B degree of realism incorporated into its dialogue. D
28. A soluble substance, alkali reacts with acids to do salts. A B C D
29. Utah is an important mining state in the Rocky Mountain regional. A B C D
30. The quality and quantity of grasses available to grazing animals may vary A B C with the season and the area. D
31. The dictionary functions primarily as a tool for the defining the meanings of words. A B C D
32. The carbohydrates, proteins, and fats in food are A B C D broken down into simpler forms in the digestive tract.
33. A letter of credit is often used to companies to finance the movement of goods A B C D between countries.
34. Each lichen consists of an alga and a fungus that lives together in a kind of A B C D partnership.
35. Magazines provide the wide variety of information and entertainment A B C D

GO ON TO THE NEXT PAGE

**BEST COPY AVAILABLE**

THIS IS THE END OF SECTION 2

**IF YOU FINISH BEFORE TIME IS CALLED, CHECK YOUR WORK  
ON SECTION 2 ONLY.  
DO NOT READ OR WORK ON ANY OTHER SECTION OF THE TEST.  
THE SUPERVISOR WILL TELL YOU WHEN TO BEGIN  
WORK ON SECTION 3.**

STOP STOP STOP STOP STOP STOP

55

53

## SECTION 3

## READING COMPREHENSION AND VOCABULARY

Time—45 minutes

This section is designed to measure your ability to understand various kinds of reading materials, as well as your ability to understand the meaning and use of words. There are two types of questions in this section, with special directions for each type.

**Directions:** In questions 1–30 each sentence has a word or phrase underlined. Below each sentence are four other words or phrases, marked (A), (B), (C), and (D). You are to choose the one word or phrase that best keeps the meaning of the original sentence if it is substituted for the underlined word or phrase. Then, on your answer sheet, find the number of the question and blacken the space that corresponds to the letter you have chosen so that the letter inside the oval cannot be seen.

**Example**

The ordinary land snail moves at the rate of about two inches per minute.

- (A) expert  
(B) active  
(C) common  
(D) colorful

**Sample Answer**

Ⓐ Ⓑ ● Ⓒ Ⓓ

The best answer is (C) because "The common land snail moves at the rate of about two inches per minute" is closest in meaning to the original sentence, "The ordinary land snail moves at the rate of about two inches per minute." Therefore, you should choose answer (C).

As soon as you understand the directions, begin work on the questions.

1. In the United States educational system, intermediate school is the transitional phase between the primary grades and high school.  
(A) stage  
(B) notion  
(C) pattern  
(D) alternative
2. Nutritionists categorize food into seven basic groups.  
(A) clarify  
(B) grind  
(C) classify  
(D) channel
3. The greatest benefit of parks and playgrounds is that they provide a safe place for children to play.  
(A) clean  
(B) pretty  
(C) distant  
(D) secure
4. A corporation is a business organization that is formed to act as a single person and is legally endowed with particular rights and duties.  
(A) by word  
(B) by law  
(C) laudably  
(D) liberally
5. The cylindrical shape of a cactus reduces moisture loss.  
(A) lessens  
(B) delays  
(C) redistributes  
(D) reverses

GO ON TO THE NEXT PAGE

6. An important part of the national government is the Foreign Service, a branch of the Department of State.  
(A) a discipline  
(B) a division  
(C) an embassy  
(D) an invasion

- (A) a discipline  
(B) a division  
(C) an embassy  
(D) an invasion

7. Urban renewal programs strive to upgrade areas that are becoming slums.  
(A) reproach  
(B) improve  
(C) fortify  
(D) uproot

- (A) reproach  
(B) improve  
(C) fortify  
(D) uproot

8. People who do not sleep enough tend to become irritable.  
(A) easily annoyed  
(B) ill  
(C) weary  
(D) stiff and sore

- (A) easily annoyed  
(B) ill  
(C) weary  
(D) stiff and sore

9. Many pure metals have little use because they are too soft, rust too easily, or have some other drawback.  
(A) property  
(B) additive  
(C) disadvantage  
(D) disparity

- (A) property  
(B) additive  
(C) disadvantage  
(D) disparity

10. Charles Goodyear's discovery of vulcanized rubber began when he inadvertently spilled a mixture of rubber and sulfur on a hot stove.  
(A) experimentally  
(B) unceasingly  
(C) incompetently  
(D) accidentally

- (A) experimentally  
(B) unceasingly  
(C) incompetently  
(D) accidentally

11. The St. Lawrence Seaway links the Great Lakes with the shipping lanes of the world.  
(A) routes  
(B) centers  
(C) stations  
(D) harbors

- (A) routes  
(B) centers  
(C) stations  
(D) harbors

12. The earliest kind of desk was a box that had a sloping lid, under which there was storage space for writing materials.  
(A) a sturdy  
(B) a polished  
(C) an inclined  
(D) an adaptable

- (A) a sturdy  
(B) a polished  
(C) an inclined  
(D) an adaptable

13. Some animals pant and sweat to speed evaporation of body moisture and thus cool themselves.  
(A) wiggle slowly  
(B) breathe quickly  
(C) rest  
(D) perspire

- (A) wiggle slowly  
(B) breathe quickly  
(C) rest  
(D) perspire

14. *The Fountainhead* and *Atlas Shrugged* by Ayn Rand are two lengthy novels that serve as vehicles for her philosophy of objectivism.  
(A) an escape from  
(B) a means of conveying  
(C) ends to  
(D) chapters in

- (A) an escape from  
(B) a means of conveying  
(C) ends to  
(D) chapters in

15. The Apollo 11 astronauts were quarantined when they returned to Earth.  
(A) praised by the press  
(B) interviewed on television  
(C) contaminated with lunar germs  
(D) isolated for medical reasons

- (A) praised by the press  
(B) interviewed on television  
(C) contaminated with lunar germs  
(D) isolated for medical reasons

16. The Faraday effect was the first demonstration of a connection between magnetism and light.  
(A) questioning  
(B) proof  
(C) achievement  
(D) symbol

- (A) questioning  
(B) proof  
(C) achievement  
(D) symbol

GO ON TO THE NEXT PAGE

18. Although most birds have only a negligible sense of smell, they have acute vision.
- (A) faulty  
(B) negative
24. A single strand of vicuña hair is four times as fine as human hair.
- (A) nice  
(B) thin  
(C) flat  
(D) wavy
- Directions:** The rest of this section is based on a variety of reading material followed by questions about the meaning of the material. For questions 31–60, you are to choose the one best answer, (A), (B), (C), or (D), to each question. Then, on your answer sheet, find

Answer all questions following a passage on the basis of what is stated or implied in that passage.

**As soon as you understand the directions, begin work on the questions.**

-16-

-17-

# Questions 31-35

Climate, more than any other single factor, determines the distribution of life on Earth. Climatic boundaries establish the limits within which organisms can survive. Plants, even more than animals, must be well adapted to climate in order to survive. They cannot move about or take shelter but must be equipped to endure whatever weather conditions are likely to occur. In the harsh conditions of the tundra, for example, low growing mosses, lichens, and a few flowering plants all hug the ground for shelter from icy winds.

Animals, despite their ability to move about and find shelter, are just as much influenced by climate as plants are. Creatures such as the camel and the penguin are so highly specialized that they have an extremely limited distribution. Others, such as bears, are flexible enough to adapt to a broad range of climates. Ocean-dwelling organisms are just as sensitive to climatic changes—in this case temperature and salinity—as land animals. Reef corals can survive only in clear warm seawater. Certain foraminifers are so sensitive to changes in their environment that their presence can be taken as an index of sea temperature. Human beings are among the least specialized of all animals and can live almost anywhere. Their clothes and their homes act as a sort of "miniature climate" that can be taken with them everywhere.

31. According to the passage, which of the following factors MOST influences where life is found on Earth?

- (A) Climatic conditions
- (B) The length of days and nights
- (C) Ocean tides
- (D) Geographic boundaries

32. According to the passage, plants on the tundra grow close to the ground

- (A) to avoid being eaten by arctic animals
- (B) because fertilizer is not readily available
- (C) to minimize exposure to the cold
- (D) because unfrozen water supplies are very scarce

33. According to the passage, which of the following can be found in the greatest number of different climate areas on Earth?

- (A) Reef corals
- (B) Penguins
- (C) Bears
- (D) Camels

34. It can be inferred from the passage that foraminifers are a

- (A) kind of weather pattern
- (B) form of sea life
- (C) species of tundra plant
- (D) type of miniature penguin

35. According to the passage, human beings can survive almost everywhere on Earth because

- (A) they have developed advanced forms of transportation
- (B) they have learned how to process seawater for drinking
- (C) their body temperature can vary considerably
- (D) their shelters and clothing help them to adapt to the environment

GO ON TO THE NEXT PAGE

# Questions 36-41

Wood carving began as a necessity in America and developed into an art. Because of the lack of other materials, early settlers were forced to make tools and utensils out of wood. At first, these articles were whittled with a knife, but when pioneer craftsmen set up their primitive shops most of them were fashioned on a lathe—a machine which holds an object and rotates it while it is being shaped by a tool.

However, even after Massachusetts-born Thomas Blanchard designed a lathe which could turn irregular shapes—an innovation that made possible mass production of gunstocks, shoe lasts, oblong and square woodenware—craftsmen who could use knife and chisel skillfully were still in demand. Some found ready employment in shops of cabinetmakers and chairmakers, while others carved decoys. Still others specialized in creating shop signs, ship figureheads, or in decorating interior woodwork. A few even accepted commissions to make busts of prominent citizens.

36. This passage most likely came from a longer work about early American

- (A) arts and crafts
- (B) political leaders
- (C) logging industries
- (D) fashion design

40. Which of the following objects was an important woodworking tool used by early American craftsmen?

- (A) A bust
- (B) A decoy
- (C) A figurehead
- (D) A chisel

37. According to the passage, the first settlers used wood for their utensils and tools because it was

- (A) durable
- (B) inexpensive
- (C) available
- (D) attractive

41. According to the passage, when wooden articles began to be mass-produced, what did many individual woodworkers do?

- (A) They moved out of Massachusetts.
- (B) They found work making specialized items.
- (C) They made demands on government leaders.
- (D) They took jobs on ships as sailors.

38. The passage suggests that the early shops for making tools were

- (A) not very sophisticated
- (B) known for doing very quick work
- (C) dependent on imported materials
- (D) frequented only by fashionable people

39. It can be inferred from the passage that, before Thomas Blanchard's invention, lathes

- (A) were not made of wood
- (B) could not produce square objects
- (C) were found only in cabinetmaking shops
- (D) could be used to make most tools and utensils

GO ON TO THE NEXT PAGE



3 • 3 • 3 • 3 • 3 • 3 • 3 • 3

3 • 3 • 3 • 3 • 3 • 3 • 3 • 3

## Questions 42-48

When a young short-toed treecreeper hatches from the egg, it is naked, blind, and helpless. At birth it weighs 0.8 gram. Its manifestations of life are wholly out of tune with its quest to achieve a weight of 8 to 9 grams as quickly as possible, the latter being the weight of an adult treecreeper. To achieve this end the fledgling must eat, and it must also be warmer than its environment. It gets food from its parents, but warmth from its mother alone. In all this the youngster does not do much of an active nature. On the first day it often stretches out of its own accord, raising its neck high and opening its beak. It also responds to the feeding calls of its parents, which are uttered only if the chick does not open up. During the first few days, the chick cannot see, but it can induce the parents to provide feedings by emitting a begging call.

Haartman (1953) took six of seven young pied flycatchers (relatives of the treecreeper) out of their nest. Since there was now only one chick in the nest, the parents were not as quick to provide frequent feedings. The other six chicks, now hungry, were then brought near the nest so that their parents could hear all seven even though they could see only one. At once the rate of attempted feedings doubled. The one chick in the nest could not handle this glut of food, naturally, so the parents ended up eating some of it themselves.

42. How many different types of birds are referred to in the passage?  
(A) One  
(B) Two  
(C) Six  
(D) Seven
43. In the passage, the author refers to a young bird as all of the following EXCEPT a  
(A) chick  
(B) fledgling  
(C) youngster  
(D) hatchling
44. According to the passage, which of the following is most important in satisfying the baby treecreeper's need for warmth?  
(A) The materials in the nest  
(B) The mother bird  
(C) The father bird  
(D) The baby bird's siblings
45. The purpose of the parent treecreeper's feeding calls is to  
(A) note the need for food  
(B) alert the babies to danger  
(C) increase the variety of food  
(D) get the baby to open its mouth
46. Which of the following helps newborn treecreepers to gain weight?  
(A) Begging noises  
(B) Low body temperature  
(C) Insulating feathers  
(D) Good vision
47. It can be inferred from the passage that the purpose of Haartman's experiment was to  
(A) study food deprivation in young birds  
(B) determine what makes birds select a nesting site  
(C) observe what stimulates parent birds to feed their young  
(D) test the eyesight of adult birds
48. It is most likely that the author is using these two paragraphs as  
(A) an introduction to a discussion of various kinds of birds  
(B) a broad description of bird development  
(C) a set of instructions for raising birds  
(D) specific examples of a preceding general statement about bird feeding

## Questions 49-54

Unlike any earlier building complex anywhere in the world, Rockefeller Center in New York City was built, not as a place where people could live, but as a city in which they could work. It was the biggest building project of its kind, a city within a city, and the forerunner of projects that have sprung up all over the world. Thirty architects, 120 draftsman, and hundreds of other artists and technicians were employed just to draft the plans. Before the buildings could be erected, 229 old buildings had to be emptied of 4,000 tenants and razed. Just to buy up the leases took over two years and cost over \$6,000,000. The unusual shape and setbacks of the seventy-story RCA building resulted primarily from practical considerations such as lighting, the movement of people, and the building's services. The lower concourse and basement level were set aside for shops. A sunken plaza, complete with gardens and fountains, was designed to provide access to these shops. Today the plaza, which is used for ice-skating in winter and dining and dancing in summer, is one of the center's most popular attractions.

49. What is the main idea of the passage?  
(A) The importance of a pleasant work environment  
(B) The purpose of the RCA building setbacks  
(C) The recreational facilities at Rockefeller Center  
(D) The architectural significance of Rockefeller Center
50. According to the passage, Rockefeller Center was originally planned to serve as what kind of complex?  
(A) Commercial  
(B) Housing  
(C) Recreational  
(D) Tourist
51. According to the passage, which of the following is true of Rockefeller Center?  
(A) It was patterned after an ancient design.  
(B) It has been imitated numerous times.  
(C) All shopkeepers were required to take two-year leases.  
(D) Four thousand tenants are located in the complex
52. According to the passage, which of the following had to be done before the actual construction of the complex could start?  
(A) Six million dollars in rent had to be collected.  
(B) Over four thousand workers had to be hired.  
(C) Over two hundred buildings had to be torn down.  
(D) The future occupants had to be put under contract.
53. According to the passage, what does the shape of the RCA building reflect?  
(A) Architectural creativity  
(B) City regulations  
(C) Practical considerations  
(D) Decreased space needs
54. The sunken plaza at Rockefeller Center was originally designed as  
(A) an entrance to shops  
(B) an ice-skating rink  
(C) an outdoor restaurant  
(D) a tourist attraction

## Questions 55-60

Minnie Marx was determined to see her sons, Leonard, Adolph, Milton, Julius, and Herbert, succeed in show business. In 1908 she organized her motley teen-agers into an act and propelled them onto the vaudeville stage. By the end of the First World War the brothers had changed their names to Chico, Harpo, Gummo, Groucho, and Zeppo, respectively; Zeppo, who had been too young to play the circuit, replaced Gummo. Their break came in 1924, when (following in the footsteps of their famous uncle, Al Shean, half of the team of Gallagher and Shean) they left the circuit to open in their own Broadway musical, *I'll Say She Is*. Its success led to even greater acclaim in the stage comedies *The Cocoanuts* and *Animal Crackers* and, when these productions were filmed by Paramount Pictures, to the Marx Brothers' debut in movies.

Sound was essential to an appreciation of the Marx Brothers' fast-paced mixture of verbal and visual gags (even the mute Harpo's honking horns and harp solos required a sound track). The brothers' contrapuntal verbal styles relied for humor on puns, aphorisms, malapropisms, wisecracks, gags, insults, and sheer nonsense. Their wit, every line of it delivered with split-second timing, made the Marx Brothers' films, from *The Cocoanuts* in 1929 to *Love Happy* in 1950, as popular as any ever shown in America and the best of them—*Monkey Business*, *Horse Feathers*, *Duck Soup*, *A Night at the Opera*, and *A Day at the Races*—unforgettable classics.

55. What is the main topic of the passage?

- (A) The role of Minnie Marx in her sons' success
- (B) The career of the Marx Brothers
- (C) The best films of the Marx Brothers
- (D) The stage names of the Marx Brothers

56. According to the passage, Julius' stage name was

- (A) Harpo
- (B) Groucho
- (C) Chico
- (D) Zeppo

57. What led to the Marx Brothers' first appearance in films?

- (A) Their success in the theater
- (B) Their mother's determination
- (C) Their uncle's influence
- (D) Their verbal style of comedy

58. The first movie that the Marx Brothers appeared in was called

- (A) *I'll Say She Is*
- (B) *The Cocoanuts*
- (C) *Monkey Business*
- (D) *Love Happy*

59. Which of the Marx Brothers did not appear in movies?

- (A) Chico
- (B) Zeppo
- (C) Harpo
- (D) Gummo

60. The author thinks which of the following about the Marx Brothers' movies?

- (A) Some are truly great.
- (B) They contain too much nonsense.
- (C) Those that rely primarily on visual humor are best.
- (D) Their popularity is undeserved.

THIS IS THE END OF SECTION 3

IF YOU FINISH BEFORE TIME IS CALLED, CHECK YOUR WORK ON SECTION 3 ONLY.  
DO NOT READ OR WORK ON ANY OTHER SECTION OF THE TEST.  
THE SUPERVISOR WILL TELL YOU WHEN TO BEGIN WORK ON SECTION 4.



GO ON TO THE NEXT PAGE



4 • 4 • 4 • 4 • 4 • 4 • 4

SECTION 4  
READING AND CONTEXT  
Time—33 minutes

**Directions:** Read the passages below. There is a number at the beginning of most of the lines in each passage. In each numbered line, there is a box with four answer choices marked (A), (B), (C), and (D). You are to choose the one answer that is most appropriate. Then, on your answer sheet, find the number of the line and blacken the space that corresponds to the letter of the answer you have chosen so that the letter inside the oval cannot be seen. Before choosing your answers, read the passage quickly for comprehension.

ExamplesSample Answer

- 1 The idea that rocks last forever and

(A) quite
(B) ever
(C) never
(D) very

change

☐ (A) ☐ (B) ☒ (C) ☐ (D)

- 2 is not completely true. If you have ever stood

(A) beside
(B) on top of
(C) underneath
(D) between

☒ (A) ☐ (B) ☐ (C) ☐ (D)

a rushing river, you have seen the water hammering away at the rocks.

In line 1, the correct answer is "never." Therefore, you should choose (C).  
In line 2, the correct answer is "beside." Therefore, you should choose (A).  
As soon as you understand the directions, begin work on the questions.

GO ON TO THE NEXT PAGE

93

4 • 4 • 4 • 4 • 4 • 4 • 4

Questions 1-16

It is generally understood that a ballad is a song

1. that tells a story.

(A) so
(B) and
(C) but
(D) as

a folk song is not

2. so

(A) easy
(B) easily
(C) ease
(D) easier

defined. A ballad is a folk song; however,

3. a folk song is not a ballad

(A) because
(B) if
(C) whether
(D) unless

it tells

4. a story. "Folk song," then, has come to

(A) be
(B) had been
(C) being
(D) been

the

5. inclusive term, covering many

(A) varieties
(B) advances
(C) conclaves
(D) adherents

of music. For instance,

6. a spiritual is a religious folk song.

(A) His
(B) The
(C) Some
(D) Its

blues,

7. a predominantly melancholy type of jazz,

(A) elected
(B) popular
(C) populous
(D) chosen

in New Orleans,

8. may

(A) qualified
(B) qualifying
(C) qualify
(D) qualifier

as folk music. In fact, there are folk

93

BEST COPY AVAILABLE

GO ON TO THE NEXT PAGE

4 • 4 • 4 • 4 • 4 • 4 • 4

4 • 4 • 4 • 4 • 4 • 4 • 4

## Questions 17-33

- 9 songs for many occupations — railroad, cattle, and so on.
10. To be (A) calculated (B) considered (C) mastered (D) treasured
11. oral (A) skills, (B) meditation, (C) transmission, (D) parallelism.
12. "Home on the Range" illustrates (A) while (B) how (C) with (D) except a folk song may
13. become a popular song. It was first (A) carried (B) tied (C) picked (D) reached up as a
14. folk song (A) at (B) in (C) to (D) for 1910 by the music collector John A. Lomax in San Antonio, Texas Lomax (A) got (B) sang (C) encountered (D) taught the song from an old man
15. had been a camp cook many years before on (A) when (B) what (C) which (D) who
16. the old Chisholm Trail By 1947 "Home on the Range" was so popular that it was adopted as the state song of Kansas
17. Tennessee is a body of (A) water (B) caverns (C) earth (D) mountain known as the Lost Sea. It
18. is listed (A) of (B) at (C) in (D) on the *Guinness Book of World Records* as
19. the world's largest underground (A) water. (B) body. (C) lake. (D) cave. The Lost Sea is part
20. of an extensive and historic cave system (A) referred to (B) known (C) indicated (D) called Craighead Caverns.
21. The caverns have been known (A) and (B) as (C) when (D) being used since the dawn
22. of the Cherokee Indian nation. (A) Which (B) That (C) One (D) Those cave expands into a
23. series of huge (A) trails (B) rooms (C) cracks (D) bodies from a small opening on the

4 • 4 • 4 • 4 • 4 • 4 • 4

24. side of the mountain. Approximately one

(A) way  
(B) road  
(C) half  
(D) mile

from the entrance.

25. in a room called "The Council Room."

(A) most  
(B) every  
(C) much  
(D) many

Indian artifacts

26. have been found.

(A) Each  
(B) Plenty  
(C) Some  
(D) Others

of the items discovered include pottery.

27. arrowheads, weapons, and jewelry. For years there were

(A) contemporary  
(B) persistent  
(C) redundant  
(D) immortal

28. rumors of a large underground

(A) deposit  
(B) cavern  
(C) jewel  
(D) lake

somewhere in a cave.

29. but it

(A) were  
(B) has  
(C) had  
(D) was

not discovered until 1905. In that year

30. a thirteen-year-old boy named Ben Sands

(A) crawled  
(B) walked  
(C) penetrated  
(D) lost

through a small

31. opening three hundred feet underground. He

(A) caught  
(B) found  
(C) saved  
(D) returned

himself in a

32. large cave filled with

(A) lake  
(B) water  
(C) bats  
(D) rumors

Today tourists visit the Lost

33. Sea and ride far out on

(A) them  
(B) tour  
(C) it  
(D) there

in glass-bottomed boats. Even though teams

of divers have tried to explore the Lost Sea, its full extent remains unknown.

### Questions 34-50

One of the recent advances in medical technology has been the development

of a new technique for examining internal tissues. Since the first brain scanner

was constructed some years ago, computed tomography, or computed

34. medical imagery,

(A) will  
(B) had  
(C) has  
(D) have

become fairly widely

35. used. Its rapid

(A) construction  
(B) acceptance  
(C) experiment  
(D) retirement

results from the fact that

36. it has overcome several of the drawbacks

(A) when  
(B) by  
(C) of  
(D) on

conventional x-ray

37. technology. To begin with,

(A) conventional  
(B) extended  
(C) retrievable  
(D) extractable

two-dimensional x-ray pictures

4 • 4 • 4 • 4 • 4 • 4 • 4

38. cannot show all of the information

- (A) containing  
(B) contained  
(C) containment  
(D) content

in a three-dimensional

39. object. Things at different depths

- (A) aren't  
(B) are  
(C) having  
(D) were

superimposed, causing confusion to

40. the viewer. Computed

- (A) demography  
(B) spectrography  
(C) photography  
(D) tomography

can give three-dimensional information.

41. The computer is able to

- (A) review  
(B) recognize  
(C) analyze  
(D) construct

pictures of the body's interior by

42. measuring the varying intensities of x-ray beams

- (A) passing  
(B) passable  
(C) pass  
(D) passive

through sections of

43. the body from hundreds of different

- (A) waves.  
(B) slides.  
(C) angles.  
(D) reports.

Such pictures are

44. based on

- (A) cereals  
(B) series  
(C) serious  
(D) serials

of contiguous cross sections, and they show

45. the

- (A) body  
(B) disease  
(C) layer  
(D) picture

as if it were separated into a sequence

46. of thin "slices." By

- (A) inspecting.  
(B) knowledge.  
(C) contrast.  
(D) medicine.

conventional x-rays generally differentiate

47. only between bone and air,

- (A) if  
(B) also  
(C) as  
(D) because

in pictures of the chest and lungs.

48. They cannot

- (A) be distinguished  
(B) distinction  
(C) distinguished  
(D) distinguish

soft tissues or variations in tissues. The

49. liver and pancreas are not

- (A) estimable  
(B) functional  
(C) dimensional  
(D) discernible

at all, and certain

50.

- (A) other  
(B) which  
(C) its  
(D) the

organs may be rendered visible only through the use of radiopaque

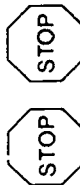
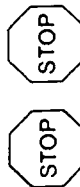
dye. Computed tomography can accurately locate a tumor and subsequently monitor

the progress of radiation treatment, so that in addition to its diagnostic capabilities, it

can play a significant role in therapy.

**THIS IS THE END OF SECTION 4**IF YOU FINISH BEFORE TIME IS CALLED, CHECK YOUR WORK  
ON SECTION 4 ONLY.

DO NOT READ OR WORK ON ANY OTHER SECTION OF THE TEST.

GO ON TO THE NEXT PAGE 

100

-31-

**BEST COPY AVAILABLE**

100

## Appendix B

### Categories of the MC Cloze Items

# Category of Each Cloze Item

<u>Item Number</u>	<u>Category</u>
1	RG
2	GR
3	RG
4	GR
5	RV
6	GR
7	RV
8	GR
9	VR
10	VR
11	RV
12	RG
13	VR
14	GR
15	VR
16	GR
17	VR
18	GR
19	RV
20	VR
21	RG
22	RG
23	RV
24	VR
25	GR
26	GR
27	VR
28	RV
29	GR
30	VR
31	VR
32	RV
33	RG
34	RG
35	RV
36	GR
37	RV
38	GR
39	RG
40	RV
41	RV
42	GR
43	RV
44	VR
45	RV
46	VR
47	RG
48	GR
49	RV
50	GR

## Appendix C

Flyer Sent to Examinees in Advance of Test Administration

Dear TOEFL Applicant:

According to our records, you are registered to take TOEFL on November 17, 1984. If you have not already received your registration confirmation ticket, it will be mailed to you soon.

In the November 1984 TOEFL, there will be four sections instead of three. The fourth section will contain a new type of question that is designed to permit comparisons with the questions in Sections 2 and 3. Your answers to the questions in Section 4 will not count toward your TOEFL score; however, it is important that you attempt to choose the best answer for each question.

The enclosed page will help you become familiar with the type of question in Section 4. Read the page carefully before you go to the test center.

We hope you find this information helpful, and we wish you every success with your academic efforts.

TOEFL Services

100



## Practice Questions For Section 4

**Directions:** Read the passages below. There is a number at the beginning of most of the lines in each passage. In each numbered line, there is a box with four answer choices marked (A), (B), (C), and (D). You are to choose the one answer that is most appropriate. Then, on your answer sheet, find the number of the line and blacken the space that corresponds to the letter of the answer you have chosen so that the letter inside the oval cannot be seen. Before choosing your answers, read the passage quickly for comprehension.

### Examples

- |   |        |  |
|---|--------|--|
| <p>1. The idea that rocks last forever and</p> <div style="border: 1px solid black; padding: 2px; display: inline-block;">             (A) quite<br/>(B) ever<br/>(C) never<br/>(D) very         </div>                         | change | <p style="text-align: right;"><u>Sample Answer</u></p> <div style="text-align: center;"> <input type="radio"/> A <input type="radio"/> B <input checked="" type="radio"/> C <input type="radio"/> D         </div> |
| <p>2. is not completely true. If you have ever stood</p> <div style="border: 1px solid black; padding: 2px; display: inline-block;">             (A) beside<br/>(B) on top of<br/>(C) underneath<br/>(D) between         </div> |        | <div style="text-align: center;"> <input type="radio"/> A <input type="radio"/> B <input type="radio"/> C <input checked="" type="radio"/> D         </div>  |

a rushing river, you have seen the water hammering away at the rocks.

In line 1, the correct answer is "never." Therefore, you should choose (C).

In line 2, the correct answer is "beside." Therefore, you should choose (A).

### Questions 1-9

The positions of the stars in the sky change slightly

1. over many years. 

(A) Besides,  
(B) However,  
(C) Consequently,  
(D) Also,

 the stars seem to
  2. sweep across the sky each night. Their 

(A) apparent  
(B) living  
(C) abrupt  
(D) stationary
  3. movement is due to the Earth's rotation. 

(A) Whenever  
(B) If  
(C) As  
(D) How
  4. the Earth rotates on its 

(A) point,  
(B) orbit,  
(C) shaft,  
(D) axis,

 we on the Earth are
  5. always moving 

(A) of  
(B) from  
(C) to  
(D) toward

 west to east. But because we
  6. do not 

(A) judge  
(B) impress  
(C) catch  
(D) sense

 this motion, it appears to us
  7. that stars revolve 

(A) up  
(B) beneath  
(C) overhead  
(D) inside

 in a westerly direction.
  8. 

(A) Only  
(B) Even  
(C) Then  
(D) Moreover

 the North Star does not seem to move because
  9. it is almost 

(A) instantly  
(B) easily  
(C) directly  
(D) northerly

 above the North Pole. The North Star
- has served as a guide to navigators since ancient times.

Questions 10-23

- Those who love music, who lose themselves in a book,  
or who can spend hours painting a picture of a barn
10. know the deep satisfaction 

(A) one
(B) it
(C) that
(D) they

 can be found in art.
11. It is not easy to 

(A) direct
(B) say
(C) feel
(D) express

 this satisfaction in words.
12. But, in 

(A) some
(B) almost
(C) somewhat
(D) any

 partly mysterious way, works of art
13. are 

(A) including
(B) among
(C) about
(D) pertaining to

 the things of highest value in our lives.
14. 

(A) That
(B) A
(C) This
(D) The

 fine piece of music, a masterful
15. 

(A) painting,
(B) drafting,
(C) depicting,
(D) designing,

 or a first-rate play has the power
16. to 

(A) release
(B) increase
(C) capture
(D) squander

 our full attention.
17. We are completely 

(A) wrapped
(B) focused
(C) looked
(D) given

 up in it,
18. and everything works out 

(A) good.
(B) order.
(C) proportion.
(D) right.

 The music
19. comes 

(A) at
(B) to
(C) in
(D) out

 the right close at the right time
20. and in the right way. The play 

(A) ends
(B) grows
(C) acts
(D) expires

 not necessarily
21. on a happy note, but in a way that 

(A) seemed
(B) seems
(C) appeared
(D) is appearing

 inevitable
22. and appropriate. As we grow more 

(A) aware
(B) understanding
(C) appreciating
(D) informed

 of the painting,
23. its parts seem to 

(A) advance
(B) live
(C) separate
(D) belong

 together. We perceive harmony

in the object and we feel harmony within ourselves.

(The correct answers are printed below, upside down.)

Correct answers: 1. B 2. A 3. C 4. D 5. B 6. D 7. C 8. A  
9. C 10. C 11. D 12. A 13. B 14. B 15. A 16. C  
17. A 18. D 19. B 20. A 21. B 22. A 23. D

## Appendix D

Varimax Factors Obtained in Two-, Three-, and  
Five-Factor Solutions for the TOEFL  
(Entries Are Factor Loadings for Each Solution)

# TWO-FACTOR VARIMAX SOLUTION

## Arabic Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.4592	0.7847
List. Comp. 2	0.3813	0.8337
List. Comp. 3	0.3141	0.8610
Struc. 1	0.6207	0.3688
Struc. 2	0.6493	0.3211
Writ. Exp. 1	0.6733	0.3225
Writ. Exp. 2	0.7292	0.2899
Writ. Exp. 3	0.7157	0.3866
Vocab. 1	0.7279	0.3825
Vocab. 2	0.7315	0.3577
Vocab. 3	0.7603	0.2691
Read. Comp. 1	0.7028	0.3404
Read. Comp. 2	0.6961	0.2721
Read. Comp. 3	0.6679	0.3166
Sum Sqd. Ldgs.	5.8267	3.2658

## Chinese Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.3168	0.8208
List. Comp. 2	0.2757	0.8328
List. Comp. 3	0.2095	0.8425
Struc. 1	0.6380	0.3352
Struc. 2	0.7057	0.2429
Writ. Exp. 1	0.5813	0.3504
Writ. Exp. 2	0.7833	0.1309
Writ. Exp. 3	0.7141	0.3082
Vocab. 1	0.6264	0.4677
Vocab. 2	0.5886	0.4835
Vocab. 3	0.5396	0.5355
Read. Comp. 1	0.5188	0.5581
Read. Comp. 2	0.4950	0.5405
Read. Comp. 3	0.4977	0.5642
Sum Sqd. Ldgs.	4.3787	4.1446

Farsi Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.4102	0.8270
List. Comp. 2	0.3858	0.8128
List. Comp. 3	0.2900	0.8664
Struc. 1	0.6164	0.4666
Struc. 2	0.6381	0.3096
Writ. Exp. 1	0.6895	0.3499
Writ. Exp. 2	0.8058	0.2348
Writ. Exp. 3	0.7129	0.3717
Vocab. 1	0.7290	0.3658
Vocab. 2	0.6902	0.4150
Vocab. 3	0.7311	0.3126
Read. Comp. 1	0.6082	0.5192
Read. Comp. 2	0.6130	0.4284
Read. Comp. 3	0.5235	0.5942
Sum Sqd. Ldgs.	5.3833	3.9343

French Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.3386	0.8516
List. Comp. 2	0.4066	0.8268
List. Comp. 3	0.3650	0.8378
Struc. 1	0.6561	0.4472
Struc. 2	0.6956	0.3461
Writ. Exp. 1	0.6196	0.4384
Writ. Exp. 2	0.7163	0.2459
Writ. Exp. 3	0.6989	0.3947
Vocab. 1	0.7887	0.3094
Vocab. 2	0.6825	0.4225
Vocab. 3	0.6371	0.4090
Read. Comp. 1	0.3947	0.7172
Read. Comp. 2	0.5839	0.5759
Read. Comp. 3	0.5096	0.6795
Sum Sqd. Ldgs.	4.9632	4.5882

Indonesian Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.7836	0.3466
List. Comp. 2	0.8141	0.2970
List. Comp. 3	0.7891	0.2436
Struc. 1	0.6594	0.2038
Struc. 2	0.5510	0.4730
Writ. Exp. 1	0.1856	0.7419
Writ. Exp. 2	0.2260	0.7933
Writ. Exp. 3	0.3619	0.6157
Vocab. 1	0.5247	0.5803
Vocab. 2	0.5836	0.4689
Vocab. 3	0.5615	0.4610
Read. Comp. 1	0.5194	0.4867
Read. Comp. 2	0.5140	0.5072
Read. Comp. 3	0.4702	0.5339
Sum Sqd. Ldgs.	4.5404	3.6401

Japanese Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.3971	0.8208
List. Comp. 2	0.3611	0.8348
List. Comp. 3	0.3078	0.8609
Struc. 1	0.7141	0.3191
Struc. 2	0.7765	0.2144
Writ. Exp. 1	0.5852	0.3879
Writ. Exp. 2	0.7428	0.2960
Writ. Exp. 3	0.7358	0.3189
Vocab. 1	0.6895	0.4460
Vocab. 2	0.7190	0.3495
Vocab. 3	0.6297	0.4444
Read. Comp. 1	0.5807	0.5526
Read. Comp. 2	0.6496	0.4655
Read. Comp. 3	0.6651	0.4200
Sum Sqd. Ldgs.	5.5218	3.8161

Korean Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.3380	0.8424
List. Comp. 2	0.3117	0.8616
List. Comp. 3	0.2381	0.8836
Struc. 1	0.7042	0.3101
Struc. 2	0.7675	0.1478
Writ. Exp. 1	0.5247	0.4767
Writ. Exp. 2	0.7641	0.2012
Writ. Exp. 3	0.7287	0.2671
Vocab. 1	0.7256	0.3274
Vocab. 2	0.7488	0.3264
Vocab. 3	0.6234	0.4623
Read. Comp. 1	0.6600	0.4557
Read. Comp. 2	0.6745	0.3855
Read. Comp. 3	0.6414	0.4605
Sum Sqd. Ldgs.	5.5210	3.6855

Spanish Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.8282	0.3464
List. Comp. 2	0.8291	0.3508
List. Comp. 3	0.8365	0.3125
Struc. 1	0.5547	0.4892
Struc. 2	0.4969	0.5245
Writ. Exp. 1	0.6661	0.4259
Writ. Exp. 2	0.5920	0.4987
Writ. Exp. 3	0.6501	0.4817
Vocab. 1	0.3339	0.7330
Vocab. 2	0.4646	0.6285
Vocab. 3	0.3377	0.6443
Read. Comp. 1	0.3445	0.7543
Read. Comp. 2	0.3183	0.7530
Read. Comp. 3	0.3577	0.7629
Sum Sqd. Ldgs.	4.6338	4.5824

Thai Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.3411	0.8222
List. Comp. 2	0.3091	0.8451
List. Comp. 3	0.2111	0.8788
Struc. 1	0.4241	0.4401
Struc. 2	0.6470	0.2719
Writ. Exp. 1	0.5376	0.3336
Writ. Exp. 2	0.7523	0.1410
Writ. Exp. 3	0.7334	0.2315
Vocab. 1	0.6571	0.3721
Vocab. 2	0.7172	0.2119
Vocab. 3	0.6503	0.3530
Read. Comp. 1	0.5800	0.4970
Read. Comp. 2	0.5847	0.4131
Read. Comp. 3	0.6010	0.4289
Sum Sqd. Ldgs.	4.6566	3.5245



## THREE-FACTOR VARIMAX SOLUTION

## Arabic Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.3914	0.3389	0.7474
List. Comp. 2	0.3318	0.2942	0.8026
List. Comp. 3	0.2660	0.2691	0.8360
Struc. 1	0.6702	0.2348	0.3091
Struc. 2	0.7022	0.2374	0.2585
Writ. Exp. 1	0.6467	0.3302	0.2625
Writ. Exp. 2	0.7387	0.3110	0.2225
Writ. Exp. 3	0.6599	0.3846	0.3244
Vocab. 1	0.5405	0.5268	0.3271
Vocab. 2	0.5193	0.5515	0.3033
Vocab. 3	0.5149	0.5880	0.2140
Read. Comp. 1	0.3465	0.6891	0.2972
Read. Comp. 2	0.2355	0.7881	0.2356
Read. Comp. 3	0.2940	0.6910	0.2776
Sum Sqd. Ldgs.	3.7646	3.2399	2.7534

## Chinese Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.3620	0.2979	0.7693
List. Comp. 2	0.3535	0.2579	0.7832
List. Comp. 3	0.2975	0.2169	0.8186
Struc. 1	0.3048	0.6130	0.2825
Struc. 2	0.2764	0.6909	0.2006
Writ. Exp. 1	0.1599	0.6369	0.3755
Writ. Exp. 2	0.3317	0.7268	0.0500
Writ. Exp. 3	0.3180	0.6877	0.2527
Vocab. 1	0.6620	0.4108	0.2306
Vocab. 2	0.6330	0.3875	0.2608
Vocab. 3	0.6133	0.3524	0.3260
Read. Comp. 1	0.6807	0.2924	0.3113
Read. Comp. 2	0.7519	0.2194	0.2482
Read. Comp. 3	0.7023	0.2564	0.3037
Sum Sqd. Ldgs.	3.4875	3.1046	2.6809

Farsi Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.3596	0.3597	0.7750
List. Comp. 2	0.3690	0.2935	0.7800
List. Comp. 3	0.2783	0.2705	0.8382
Struc. 1	0.4932	0.4640	0.3753
Struc. 2	0.4200	0.5981	0.1688
Writ. Exp. 1	0.7169	0.2063	0.3397
Writ. Exp. 2	0.7507	0.3530	0.1785
Writ. Exp. 3	0.7110	0.2662	0.3447
Vocab. 1	0.6389	0.4179	0.2908
Vocab. 2	0.6264	0.3731	0.3545
Vocab. 3	0.6853	0.3331	0.2623
Read. Comp. 1	0.3681	0.6694	0.3640
Read. Comp. 2	0.3194	0.7425	0.2456
Read. Comp. 3	0.2604	0.7009	0.4278
Sum Sqd. Ldgs.	3.9210	3.0151	3.0019

French Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.8077	0.3955	0.1895
List. Comp. 2	0.7777	0.4295	0.2490
List. Comp. 3	0.7889	0.4383	0.1793
Struc. 1	0.3604	0.7265	0.2354
Struc. 2	0.2683	0.6238	0.3928
Writ. Exp. 1	0.3564	0.6872	0.2235
Writ. Exp. 2	0.2029	0.2715	0.7986
Writ. Exp. 3	0.3310	0.4871	0.5561
Vocab. 1	0.2347	0.5697	0.5854
Vocab. 2	0.3427	0.6500	0.3557
Vocab. 3	0.3256	0.6948	0.2366
Read. Comp. 1	0.6909	0.2032	0.4656
Read. Comp. 2	0.5288	0.3582	0.5544
Read. Comp. 3	0.6426	0.2791	0.5467
Sum Sqd. Ldgs.	3.8076	3.7106	2.6848

Indonesian Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.7153	0.4277	0.2266
List. Comp. 2	0.7670	0.3869	0.1921
List. Comp. 3	0.7946	0.2815	0.1799
Struc. 1	0.3815	0.6709	-0.0456
Struc. 2	0.3518	0.6090	0.2717
Writ. Exp. 1	0.3033	0.1111	0.7721
Writ. Exp. 2	0.1596	0.4228	0.6901
Writ. Exp. 3	0.1831	0.5739	0.4344
Vocab. 1	0.4908	0.3824	0.4899
Vocab. 2	0.5717	0.3275	0.3956
Vocab. 3	0.6218	0.2076	0.4398
Read. Comp. 1	0.2682	0.6853	0.2508
Read. Comp. 2	0.2955	0.6401	0.2934
Read. Comp. 3	0.2733	0.6037	0.3367
Sum Sqd. Ldgs.	3.3401	3.2938	2.3099

Japanese Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.3776	0.7751	0.3039
List. Comp. 2	0.3473	0.7942	0.2859
List. Comp. 3	0.3099	0.8265	0.2525
Struc. 1	0.5249	0.2454	0.5279
Struc. 2	0.5045	0.1452	0.6223
Writ. Exp. 1	0.1339	0.3906	0.7599
Writ. Exp. 2	0.4318	0.2420	0.6621
Writ. Exp. 3	0.4444	0.2625	0.6425
Vocab. 1	0.6631	0.3456	0.3685
Vocab. 2	0.6988	0.2411	0.3588
Vocab. 3	0.6951	0.3352	0.2494
Read. Comp. 1	0.6550	0.4519	0.2375
Read. Comp. 2	0.6937	0.3578	0.2826
Read. Comp. 3	0.6795	0.3148	0.3125
Sum Sqd. Ldgs.	4.0742	2.9973	2.8929

Korean Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.3310	0.2594	0.8070
List. Comp. 2	0.2987	0.2539	0.8333
List. Comp. 3	0.2806	0.1768	0.8552
Struc. 1	0.3678	0.6387	0.2760
Struc. 2	0.3431	0.7214	0.1197
Writ. Exp. 1	0.1584	0.5959	0.4887
Writ. Exp. 2	0.3218	0.7400	0.1800
Writ. Exp. 3	0.3048	0.7157	0.2497
Vocab. 1	0.5696	0.5061	0.2431
Vocab. 2	0.6493	0.4716	0.2230
Vocab. 3	0.5973	0.3683	0.3674
Read. Comp. 1	0.6967	0.3345	0.3375
Read. Comp. 2	0.7632	0.2918	0.2490
Read. Comp. 3	0.7286	0.2857	0.3328
Sum Sqd. Ldgs.	3.4535	3.3993	3.0317

Spanish Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.3388	0.3626	0.7778
List. Comp. 2	0.3494	0.3494	0.7875
List. Comp. 3	0.3206	0.3231	0.8090
Struc. 1	0.3224	0.6630	0.2832
Struc. 2	0.3371	0.6975	0.1971
Writ. Exp. 1	0.2711	0.6614	0.4092
Writ. Exp. 2	0.3078	0.7327	0.2846
Writ. Exp. 3	0.3368	0.6436	0.4065
Vocab. 1	0.6505	0.4388	0.1850
Vocab. 2	0.5810	0.3830	0.3636
Vocab. 3	0.6764	0.1594	0.3548
Read. Comp. 1	0.7043	0.3713	0.2411
Read. Comp. 2	0.7421	0.2722	0.2719
Read. Comp. 3	0.7129	0.3773	0.2537
Sum Sqd. Ldgs.	3.6121	3.3916	2.9011

Thai Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.2574	0.7958	0.3094
List. Comp. 2	0.2685	0.8224	0.2478
List. Comp. 3	0.2224	0.8639	0.1532
Struc. 1	0.3253	0.4099	0.3170
Struc. 2	0.6709	0.2359	0.2275
Writ. Exp. 1	0.5947	0.3050	0.1548
Writ. Exp. 2	0.7857	0.1001	0.2393
Writ. Exp. 3	0.6360	0.1850	0.4025
Vocab. 1	0.2790	0.3157	0.7344
Vocab. 2	0.2970	0.1509	0.7903
Vocab. 3	0.2803	0.2975	0.7201
Read. Comp. 1	0.5278	0.4601	0.3209
Read. Comp. 2	0.5991	0.3795	0.2321
Read. Comp. 3	0.4987	0.3887	0.3825
Sum Sqd. Ldgs.	3.2499	3.1345	2.5615

# FIVE-FACTOR VARIMAX SOLUTION

## Arabic Group

	<u>Factor</u>				
<u>Parcel</u>	1	2	3	4	5
List. Comp. 1	0.7390	0.2629	0.2852	0.2806	0.2321
List. Comp. 2	0.7946	0.2307	0.2498	0.2224	0.2142
List. Comp. 3	0.8265	0.2439	0.1666	0.1832	0.1966
Struc. 1	0.2708	0.2648	0.1646	0.2417	0.7292
Struc. 2	0.2259	0.1836	0.3290	0.2103	0.7177
Writ. Exp. 1	0.2657	0.2180	0.2850	0.7650	0.1138
Writ. Exp. 2	0.2091	0.2655	0.2123	0.6934	0.3610
Writ. Exp. 3	0.3107	0.3158	0.2846	0.5627	0.3493
Vocab. 1	0.3215	0.3010	0.6331	0.2838	0.2825
Vocab. 2	0.3903	0.3128	0.6527	0.2908	0.2359
Vocab. 3	0.2135	0.3005	0.7567	0.2442	0.2227
Read. Comp. 1	0.2783	0.6663	0.2537	0.3009	0.2068
Read. Comp. 2	0.2183	0.7453	0.3073	0.1910	0.1351
Read. Comp. 3	0.2527	0.7048	0.2016	0.1973	0.2603
Sum Sqd. Ldgs.	2.6240	2.2785	2.1130	2.0378	1.7642

## Chinese Group

	<u>Factor</u>				
<u>Parcel</u>	1	2	3	4	5
List. Comp. 1	0.7643	0.2421	0.2706	0.2652	0.1666
List. Comp. 2	0.7835	0.2213	0.2767	0.2342	0.1301
List. Comp. 3	0.8196	0.1819	0.2264	0.2060	0.1189
Struc. 1	0.3107	0.6031	0.1044	0.3687	0.0711
Struc. 2	0.2416	0.7673	0.2745	0.0730	0.0599
Writ. Exp. 1	0.2485	0.2827	0.1901	0.1924	0.8683
Writ. Exp. 2	0.0529	0.6876	0.2496	0.2480	0.2271
Writ. Exp. 3	0.2507	0.6150	0.1782	0.3286	0.2475
Vocab. 1	0.2291	0.3350	0.3820	0.6193	0.1004
Vocab. 2	0.2476	0.2620	0.2898	0.7106	0.1442
Vocab. 3	0.3118	0.2234	0.2701	0.7043	0.1429
Read. Comp. 1	0.2892	0.2447	0.6637	0.2822	0.1965
Read. Comp. 2	0.2390	0.2240	0.7665	0.2420	0.1020
Read. Comp. 3	0.3025	0.2624	0.6725	0.2799	0.0794
Sum Sqd. Ldgs.	2.5953	2.4337	2.2008	2.1166	1.0408

Farsi Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.7658	0.2574	0.3056	0.2826	0.1909
List. Comp. 2	0.7678	0.2592	0.3217	0.2340	0.1361
List. Comp. 3	0.8297	0.2593	0.1779	0.2391	0.1150
Struc. 1	0.3655	0.4446	0.2572	0.3734	0.2872
Struc. 2	0.2076	0.2800	0.2269	0.2119	0.8589
Writ. Exp. 1	0.3247	0.7271	0.2397	0.1448	0.2079
Writ. Exp. 2	0.1523	0.7337	0.3086	0.3258	0.1686
Writ. Exp. 3	0.3203	0.6810	0.3153	0.2399	0.1324
Vocab. 1	0.2680	0.2765	0.7126	0.2747	0.2290
Vocab. 2	0.3296	0.2707	0.7077	0.2503	0.1789
Vocab. 3	0.2286	0.3278	0.7400	0.2508	0.0931
Read. Comp. 1	0.3402	0.3324	0.2535	0.6662	0.1664
Read. Comp. 2	0.2132	0.2511	0.3002	0.7785	0.0908
Read. Comp. 3	0.4225	0.1704	0.2525	0.5947	0.3427
Sum Sqd. Ldgs.	2.8446	2.4674	2.3726	2.1653	1.2240

French Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.7987	0.2520	0.2874	0.1819	0.2086
List. Comp. 2	0.7620	0.2939	0.3112	0.2305	0.2188
List. Comp. 3	0.7852	0.2898	0.2756	0.1706	0.2324
Struc. 1	0.3740	0.4886	0.2049	0.1462	0.5515
Struc. 2	0.2784	0.1934	0.2368	0.2597	0.8165
Writ. Exp. 1	0.4768	0.5396	-0.0380	0.3990	0.2781
Writ. Exp. 2	0.1937	0.1469	0.2780	0.8448	0.2019
Writ. Exp. 3	0.3293	0.4139	0.2549	0.5781	0.1958
Vocab. 1	0.1013	0.5076	0.5272	0.3090	0.3775
Vocab. 2	0.2932	0.6165	0.3288	0.2323	0.2690
Vocab. 3	0.2728	0.7842	0.2981	0.1208	0.1150
Read. Comp. 1	0.5249	0.1711	0.6197	0.2215	0.1354
Read. Comp. 2	0.3310	0.3418	0.6879	0.2186	0.2354
Read. Comp. 3	0.4738	0.2420	0.6424	0.2901	0.1801
Sum Sqd. Ldgs.	3.2061	2.4513	2.2420	1.7563	1.6070

# Indonesian Group

## Factor

<u>Parcel</u>	1	2	3	4	5
List. Comp. 1	0.6908	0.3050	0.2178	0.3244	0.2386
List. Comp. 2	0.7293	0.3070	0.2654	0.2574	0.1718
List. Comp. 3	0.7489	0.2057	0.2850	0.2085	0.1607
Struc. 1	0.3466	0.1756	0.0835	0.7881	0.0194
Struc. 2	0.2960	0.3109	0.2029	0.5719	0.2691
Writ. Exp. 1	0.3072	0.1484	0.1361	0.0709	0.8478
Writ. Exp. 2	0.0675	0.3244	0.3332	0.3045	0.6170
Writ. Exp. 3	0.0237	0.2609	0.4454	0.5556	0.2981
Vocab. 1	0.3057	0.1891	0.5913	0.3551	0.3156
Vocab. 2	0.3464	0.2611	0.7265	0.1813	0.1116
Vocab. 3	0.4624	0.2559	0.5913	0.0335	0.2324
Read. Comp. 1	0.2867	0.7475	0.0818	0.2349	0.1978
Read. Comp. 2	0.2619	0.7222	0.2263	0.1856	0.1671
Read. Comp. 3	0.1760	0.6700	0.3835	0.1739	0.1304
Sum Sqd. Ldgs.	2.4863	2.2511	2.0072	1.8495	1.6234

# Japanese Group

## Factor

<u>Parcel</u>	1	2	3	4	5
List. Comp. 1	0.7711	0.2952	0.2493	0.2845	0.1578
List. Comp. 2	0.7861	0.2480	0.2694	0.2451	0.1760
List. Comp. 3	0.8175	0.2129	0.2286	0.2416	0.1658
Struc. 1	0.2942	0.6933	0.2352	0.2994	0.0608
Struc. 2	0.1738	0.7241	0.2070	0.3217	0.1724
Writ. Exp. 1	0.2660	0.2743	0.2081	0.2053	0.8505
Writ. Exp. 2	0.2763	0.6799	0.3222	0.1433	0.2486
Writ. Exp. 3	0.2464	0.5522	0.3522	0.2330	0.3703
Vocab. 1	0.3373	0.3764	0.6624	0.2467	0.1421
Vocab. 2	0.2186	0.3490	0.7081	0.2693	0.1590
Vocab. 3	0.2902	0.2032	0.7332	0.3041	0.1590
Read. Comp. 1	0.3583	0.2053	0.3129	0.6943	0.2141
Read. Comp. 2	0.2956	0.3380	0.2998	0.6773	0.1348
Read. Comp. 3	0.2611	0.3865	0.2607	0.6702	0.1272
Sum Sqd. Ldgs.	2.7351	2.6519	2.2898	2.1252	1.1899



Korean Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.8035	0.2306	0.2348	0.2528	0.1482
List. Comp. 2	0.8302	0.2281	0.2146	0.2242	0.1526
List. Comp. 3	0.8498	0.1591	0.2364	0.1608	0.1499
Struc. 1	0.2975	0.6525	0.2265	0.3038	0.0890
Struc. 2	0.1648	0.8058	0.2636	0.1732	0.0215
Writ. Exp. 1	0.3163	0.2641	0.2048	0.2182	0.8263
Writ. Exp. 2	0.1386	0.6463	0.2466	0.2767	0.3388
Writ. Exp. 3	0.2266	0.6517	0.2114	0.2738	0.2749
Vocab. 1	0.2113	0.3644	0.2675	0.6764	0.1945
Vocab. 2	0.2032	0.3614	0.3533	0.6747	0.1362
Vocab. 3	0.3468	0.2336	0.2493	0.7364	0.1235
Read. Comp. 1	0.3239	0.3409	0.6874	0.2477	0.1345
Read. Comp. 2	0.2314	0.2991	0.7673	0.2527	0.1232
Read. Comp. 3	0.3041	0.2492	0.6809	0.3345	0.1646
Sum Sqd. Ldgs.	2.8003	2.7035	2.2088	2.1535	1.0812

Spanish Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.7784	0.3035	0.2927	0.1810	0.2094
List. Comp. 2	0.7821	0.2999	0.2814	0.2198	0.2011
List. Comp. 3	0.8054	0.3135	0.2835	0.1625	0.1235
Struc. 1	0.2667	0.5190	0.2132	0.2439	0.4574
Struc. 2	0.2356	0.2930	0.2953	0.1479	0.8044
Writ. Exp. 1	0.3804	0.6757	0.2118	0.1465	0.2131
Writ. Exp. 2	0.2745	0.7117	0.3144	0.0442	0.2569
Writ. Exp. 3	0.3770	0.6601	0.2678	0.1817	0.2073
Vocab. 1	0.1376	0.5407	0.5152	0.3778	0.0459
Vocab. 2	0.3161	0.4673	0.4257	0.4011	0.0615
Vocab. 3	0.2669	0.1757	0.2723	0.8435	0.1922
Read. Comp. 1	0.2695	0.2753	0.7390	0.1375	0.2251
Read. Comp. 2	0.2939	0.2141	0.7566	0.1861	0.1477
Read. Comp. 3	0.2738	0.2737	0.7083	0.2035	0.2549
Sum Sqd. Ldgs.	2.7791	2.7725	2.7343	1.3554	1.2905

Thai Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.7795	0.2961	0.2602	0.1758	0.1195
List. Comp. 2	0.8076	0.2319	0.2293	0.2005	0.1607
List. Comp. 3	0.8488	0.1375	0.1949	0.1612	0.1767
Struc. 1	0.2806	0.2513	0.1102	0.1591	0.8275
Struc. 2	0.3036	0.2152	0.1243	0.7794	0.0345
Writ. Exp. 1	0.1362	0.0986	0.6375	0.1865	0.4768
Writ. Exp. 2	0.1056	0.2103	0.2843	0.7595	0.1782
Writ. Exp. 3	0.1521	0.3735	0.3650	0.5207	0.1919
Vocab. 1	0.3122	0.7226	0.1821	0.2534	0.1036
Vocab. 2	0.1387	0.7732	0.1462	0.2757	0.1726
Vocab. 3	0.2427	0.7015	0.3410	0.1250	0.1420
Read. Comp. 1	0.4019	0.2980	0.5170	0.3107	0.0984
Read. Comp. 2	0.2996	0.2056	0.6184	0.3166	0.0941
Read. Comp. 3	0.3006	0.3654	0.7092	0.1651	-0.0372
Sum Sqd. Ldgs.	2.7213	2.3400	2.1298	1.9925	1.1338

## Appendix E

### Confirmatory Factor Analyses: Factor Loadings for One-, Two-, Three-, and Five-Factor Solutions

# ONE-FACTOR SOLUTION

## Arabic Group

<u>Parcel</u>	<u>Factor</u>
	1
List. Comp. 1	0.815
List. Comp. 2	0.777
List. Comp. 3	0.733
Struc. 1	0.692
Struc. 2	0.688
Writ. Exp. 1	0.711
Writ. Exp. 2	0.740
Writ. Exp. 3	0.791
Vocab. 1	0.801
Vocab. 2	0.789
Vocab. 3	0.758
Read. Comp. 1	0.747
Read. Comp. 2	0.699
Read. Comp. 3	0.702

## Chinese Group

<u>Parcel</u>	<u>Factor</u>
	1
List. Comp. 1	0.790
List. Comp. 2	0.768
List. Comp. 3	0.725
Struc. 1	0.656
Struc. 2	0.637
Writ. Exp. 1	0.625
Writ. Exp. 2	0.612
Writ. Exp. 3	0.692
Vocab. 1	0.753
Vocab. 2	0.736
Vocab. 3	0.740
Read. Comp. 1	0.739
Read. Comp. 2	0.707
Read. Comp. 3	0.728

Farsi Group

<u>Parcel</u>	<u>Factor</u>
	1
List. Comp. 1	0.841
List. Comp. 2	0.810
List. Comp. 3	0.769
Struc. 1	0.749
Struc. 2	0.656
Writ. Exp. 1	0.727
Writ. Exp. 2	0.741
Writ. Exp. 3	0.763
Vocab. 1	0.774
Vocab. 2	0.777
Vocab. 3	0.737
Read. Comp. 1	0.780
Read. Comp. 2	0.719
Read. Comp. 3	0.763

French Group

<u>Parcel</u>	<u>Factor</u>
	1
List. Comp. 1	0.843
List. Comp. 2	0.876
List. Comp. 3	0.851
Struc. 1	0.754
Struc. 2	0.706
Writ. Exp. 1	0.718
Writ. Exp. 2	0.647
Writ. Exp. 3	0.745
Vocab. 1	0.744
Vocab. 2	0.755
Vocab. 3	0.710
Read. Comp. 1	0.768
Read. Comp. 2	0.800
Read. Comp. 3	0.829

Indonesian Group

<u>Parcel</u>	<u>Factor</u>
	1
List. Comp. 1	0.817
List. Comp. 2	0.809
List. Comp. 3	0.748
Struc. 1	0.598
Struc. 2	0.694
Writ. Exp. 1	0.588
Writ. Exp. 2	0.650
Writ. Exp. 3	0.638
Vocab. 1	0.752
Vocab. 2	0.725
Vocab. 3	0.700
Read. Comp. 1	0.678
Read. Comp. 2	0.689
Read. Comp. 3	0.667

Japanese Group

<u>Parcel</u>	<u>Factor</u>
	1
List. Comp. 1	0.813
List. Comp. 2	0.792
List. Comp. 3	0.766
Struc. 1	0.731
Struc. 2	0.711
Writ. Exp. 1	0.669
Writ. Exp. 2	0.738
Writ. Exp. 3	0.749
Vocab. 1	0.802
Vocab. 2	0.759
Vocab. 3	0.749
Read. Comp. 1	0.783
Read. Comp. 2	0.781
Read. Comp. 3	0.762

Korean Group

<u>Parcel</u>	<u>Factor</u>
	1
List. Comp. 1	0.758
List. Comp. 2	0.748
List. Comp. 3	0.702
Struc. 1	0.722
Struc. 2	0.670
Writ. Exp. 1	0.676
Writ. Exp. 2	0.700
Writ. Exp. 3	0.713
Vocab. 1	0.756
Vocab. 2	0.778
Vocab. 3	0.759
Read. Comp. 1	0.787
Read. Comp. 2	0.754
Read. Comp. 3	0.775

Spanish Group

<u>Parcel</u>	<u>Factor</u>
	1
List. Comp. 1	0.836
List. Comp. 2	0.840
List. Comp. 3	0.818
Struc. 1	0.709
Struc. 2	0.691
Writ. Exp. 1	0.751
Writ. Exp. 2	0.746
Writ. Exp. 3	0.782
Vocab. 1	0.719
Vocab. 2	0.746
Vocab. 3	0.661
Read. Comp. 1	0.746
Read. Comp. 2	0.726
Read. Comp. 3	0.762

Thai Group

<u>Parcel</u>	<u>Factor</u>
	1
List. Comp. 1	0.785
List. Comp. 2	0.775
List. Comp. 3	0.719
Struc. 1	0.569
Struc. 2	0.633
Writ. Exp. 1	0.586
Writ. Exp. 2	0.624
Writ. Exp. 3	0.672
Vocab. 1	0.715
Vocab. 2	0.649
Vocab. 3	0.694
Read. Comp. 1	0.741
Read. Comp. 2	0.682
Read. Comp. 3	0.710



# TWO-FACTOR SOLUTION

## Arabic Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.897	0.0
List. Comp. 2	0.871	0.0
List. Comp. 3	0.830	0.0
Struc. 1	0.0	0.695
Struc. 2	0.0	0.693
Writ. Exp. 1	0.0	0.717
Writ. Exp. 2	0.0	0.751
Writ. Exp. 3	0.0	0.797
Vocab. 1	0.0	0.809
Vocab. 2	0.0	0.798
Vocab. 3	0.0	0.772
Read. Comp. 1	0.0	0.753
Read. Comp. 2	0.0	0.707
Read. Comp. 3	0.0	0.707

## Chinese Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.880	0.0
List. Comp. 2	0.860	0.0
List. Comp. 3	0.816	0.0
Struc. 1	0.0	0.665
Struc. 2	0.0	0.651
Writ. Exp. 1	0.0	0.627
Writ. Exp. 2	0.0	0.635
Writ. Exp. 3	0.0	0.705
Vocab. 1	0.0	0.770
Vocab. 2	0.0	0.748
Vocab. 3	0.0	0.746
Read. Comp. 1	0.0	0.746
Read. Comp. 2	0.0	0.715
Read. Comp. 3	0.0	0.733

# Farsi Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.917	0.0
List. Comp. 2	0.878	0.0
List. Comp. 3	0.854	0.0
Struc. 1	0.0	0.753
Struc. 2	0.0	0.666
Writ. Exp. 1	0.0	0.736
Writ. Exp. 2	0.0	0.763
Writ. Exp. 3	0.0	0.773
Vocab. 1	0.0	0.788
Vocab. 2	0.0	0.786
Vocab. 3	0.0	0.752
Read. Comp. 1	0.0	0.784
Read. Comp. 2	0.0	0.728
Read. Comp. 3	0.0	0.760

# French Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.906	0.0
List. Comp. 2	0.923	0.0
List. Comp. 3	0.903	0.0
Struc. 1	0.0	0.765
Struc. 2	0.0	0.719
Writ. Exp. 1	0.0	0.723
Writ. Exp. 2	0.0	0.669
Writ. Exp. 3	0.0	0.760
Vocab. 1	0.0	0.780
Vocab. 2	0.0	0.770
Vocab. 3	0.0	0.724
Read. Comp. 1	0.0	0.765
Read. Comp. 2	0.0	0.817
Read. Comp. 3	0.0	0.829

Indonesian Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.865	0.0
List. Comp. 2	0.864	0.0
List. Comp. 3	0.790	0.0
Struc. 1	0.0	0.598
Struc. 2	0.0	0.703
Writ. Exp. 1	0.0	0.598
Writ. Exp. 2	0.0	0.670
Writ. Exp. 3	0.0	0.652
Vocab. 1	0.0	0.763
Vocab. 2	0.0	0.732
Vocab. 3	0.0	0.702
Read. Comp. 1	0.0	0.687
Read. Comp. 2	0.0	0.698
Read. Comp. 3	0.0	0.683

Japanese Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.898	0.0
List. Comp. 2	0.881	0.0
List. Comp. 3	0.858	0.0
Struc. 1	0.0	0.742
Struc. 2	0.0	0.727
Writ. Exp. 1	0.0	0.669
Writ. Exp. 2	0.0	0.750
Writ. Exp. 3	0.0	0.760
Vocab. 1	0.0	0.809
Vocab. 2	0.0	0.773
Vocab. 3	0.0	0.755
Read. Comp. 1	0.0	0.781
Read. Comp. 2	0.0	0.787
Read. Comp. 3	0.0	0.770

# Korean Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.886	0.0
List. Comp. 2	0.900	0.0
List. Comp. 3	0.860	0.0
Struc. 1	0.0	0.731
Struc. 2	0.0	0.690
Writ. Exp. 1	0.0	0.666
Writ. Exp. 2	0.0	0.719
Writ. Exp. 3	0.0	0.726
Vocab. 1	0.0	0.770
Vocab. 2	0.0	0.796
Vocab. 3	0.0	0.758
Read. Comp. 1	0.0	0.790
Read. Comp. 2	0.0	0.763
Read. Comp. 3	0.0	0.777

# Spanish Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.900	0.0
List. Comp. 2	0.907	0.0
List. Comp. 3	0.886	0.0
Struc. 1	0.0	0.718
Struc. 2	0.0	0.703
Writ. Exp. 1	0.0	0.752
Writ. Exp. 2	0.0	0.757
Writ. Exp. 3	0.0	0.785
Vocab. 1	0.0	0.742
Vocab. 2	0.0	0.756
Vocab. 3	0.0	0.668
Read. Comp. 1	0.0	0.766
Read. Comp. 2	0.0	0.742
Read. Comp. 3	0.0	0.783

Thai Group

<u>Parcel</u>	<u>Factor</u>	
	1	2
List. Comp. 1	0.875	0.0
List. Comp. 2	0.881	0.0
List. Comp. 3	0.833	0.0
Struc. 1	0.0	0.567
Struc. 2	0.0	0.648
Writ. Exp. 1	0.0	0.593
Writ. Exp. 2	0.0	0.655
Writ. Exp. 3	0.0	0.701
Vocab. 1	0.0	0.729
Vocab. 2	0.0	0.678
Vocab. 3	0.0	0.711
Read. Comp. 1	0.0	0.745
Read. Comp. 2	0.0	0.694
Read. Comp. 3	0.0	0.722

# THREE-FACTOR SOLUTION

## Arabic Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.897	0.0	0.0
List. Comp. 2	0.871	0.0	0.0
List. Comp. 3	0.830	0.0	0.0
Struc. 1	0.0	0.709	0.0
Struc. 2	0.0	0.704	0.0
Writ. Exp. 1	0.0	0.731	0.0
Writ. Exp. 2	0.0	0.777	0.0
Writ. Exp. 3	0.0	0.819	0.0
Vocab. 1	0.0	0.0	0.818
Vocab. 2	0.0	0.0	0.808
Vocab. 3	0.0	0.0	0.784
Read. Comp. 1	0.0	0.0	0.760
Read. Comp. 2	0.0	0.0	0.720
Read. Comp. 3	0.0	0.0	0.715

## Chinese Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.880	0.0	0.0
List. Comp. 2	0.861	0.0	0.0
List. Comp. 3	0.816	0.0	0.0
Struc. 1	0.0	0.695	0.0
Struc. 2	0.0	0.692	0.0
Writ. Exp. 1	0.0	0.657	0.0
Writ. Exp. 2	0.0	0.678	0.0
Writ. Exp. 3	0.0	0.752	0.0
Vocab. 1	0.0	0.0	0.776
Vocab. 2	0.0	0.0	0.755
Vocab. 3	0.0	0.0	0.754
Read. Comp. 1	0.0	0.0	0.756
Read. Comp. 2	0.0	0.0	0.728
Read. Comp. 3	0.0	0.0	0.745

# Farsi Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.917	0.0	0.0
List. Comp. 2	0.878	0.0	0.0
List. Comp. 3	0.853	0.0	0.0
Struc. 1	0.0	0.762	0.0
Struc. 2	0.0	0.676	0.0
Writ. Exp. 1	0.0	0.761	0.0
Writ. Exp. 2	0.0	0.791	0.0
Writ. Exp. 3	0.0	0.797	0.0
Vocab. 1	0.0	0.0	0.797
Vocab. 2	0.0	0.0	0.795
Vocab. 3	0.0	0.0	0.758
Read. Comp. 1	0.0	0.0	0.791
Read. Comp. 2	0.0	0.0	0.737
Read. Comp. 3	0.0	0.0	0.768

# French Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.906	0.0	0.0
List. Comp. 2	0.923	0.0	0.0
List. Comp. 3	0.903	0.0	0.0
Struc. 1	0.0	0.784	0.0
Struc. 2	0.0	0.741	0.0
Writ. Exp. 1	0.0	0.746	0.0
Writ. Exp. 2	0.0	0.683	0.0
Writ. Exp. 3	0.0	0.779	0.0
Vocab. 1	0.0	0.0	0.779
Vocab. 2	0.0	0.0	0.768
Vocab. 3	0.0	0.0	0.723
Read. Comp. 1	0.0	0.0	0.777
Read. Comp. 2	0.0	0.0	0.830
Read. Comp. 3	0.0	0.0	0.844

Indonesian Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.864	0.0	0.0
List. Comp. 2	0.865	0.0	0.0
List. Comp. 3	0.791	0.0	0.0
Struc. 1	0.0	0.607	0.0
Struc. 2	0.0	0.720	0.0
Writ. Exp. 1	0.0	0.610	0.0
Writ. Exp. 2	0.0	0.689	0.0
Writ. Exp. 3	0.0	0.669	0.0
Vocab. 1	0.0	0.0	0.763
Vocab. 2	0.0	0.0	0.738
Vocab. 3	0.0	0.0	0.708
Read. Comp. 1	0.0	0.0	0.689
Read. Comp. 2	0.0	0.0	0.701
Read. Comp. 3	0.0	0.0	0.686

Japanese Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.898	0.0	0.0
List. Comp. 2	0.881	0.0	0.0
List. Comp. 3	0.858	0.0	0.0
Struc. 1	0.0	0.758	0.0
Struc. 2	0.0	0.753	0.0
Writ. Exp. 1	0.0	0.688	0.0
Writ. Exp. 2	0.0	0.777	0.0
Writ. Exp. 3	0.0	0.783	0.0
Vocab. 1	0.0	0.0	0.814
Vocab. 2	0.0	0.0	0.777
Vocab. 3	0.0	0.0	0.764
Read. Comp. 1	0.0	0.0	0.791
Read. Comp. 2	0.0	0.0	0.794
Read. Comp. 3	0.0	0.0	0.774



Korean Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.886	0.0	0.0
List. Comp. 2	0.900	0.0	0.0
List. Comp. 3	0.860	0.0	0.0
Struc. 1	0.0	0.757	0.0
Struc. 2	0.0	0.725	0.0
Writ. Exp. 1	0.0	0.679	0.0
Writ. Exp. 2	0.0	0.755	0.0
Writ. Exp. 3	0.0	0.761	0.0
Vocab. 1	0.0	0.0	0.771
Vocab. 2	0.0	0.0	0.800
Vocab. 3	0.0	0.0	0.766
Read. Comp. 1	0.0	0.0	0.800
Read. Comp. 2	0.0	0.0	0.774
Read. Comp. 3	0.0	0.0	0.790

Spanish Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.900	0.0	0.0
List. Comp. 2	0.907	0.0	0.0
List. Comp. 3	0.886	0.0	0.0
Struc. 1	0.0	0.735	0.0
Struc. 2	0.0	0.714	0.0
Writ. Exp. 1	0.0	0.780	0.0
Writ. Exp. 2	0.0	0.783	0.0
Writ. Exp. 3	0.0	0.810	0.0
Vocab. 1	0.0	0.0	0.756
Vocab. 2	0.0	0.0	0.763
Vocab. 3	0.0	0.0	0.683
Read. Comp. 1	0.0	0.0	0.791
Read. Comp. 2	0.0	0.0	0.771
Read. Comp. 3	0.0	0.0	0.810

Thai Group

<u>Parcel</u>	<u>Factor</u>		
	1	2	3
List. Comp. 1	0.876	0.0	0.0
List. Comp. 2	0.880	0.0	0.0
List. Comp. 3	0.832	0.0	0.0
Struc. 1	0.0	0.570	0.0
Struc. 2	0.0	0.672	0.0
Writ. Exp. 1	0.0	0.611	0.0
Writ. Exp. 2	0.0	0.695	0.0
Writ. Exp. 3	0.0	0.733	0.0
Vocab. 1	0.0	0.0	0.736
Vocab. 2	0.0	0.0	0.681
Vocab. 3	0.0	0.0	0.718
Read. Comp. 1	0.0	0.0	0.751
Read. Comp. 2	0.0	0.0	0.697
Read. Comp. 3	0.0	0.0	0.730

# FIVE-FACTOR SOLUTION

## Arabic Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.897	0.0	0.0	0.0	0.0
List. Comp. 2	0.871	0.0	0.0	0.0	0.0
List. Comp. 3	0.830	0.0	0.0	0.0	0.0
Struc. 1	0.0	0.731	0.0	0.0	0.0
Struc. 2	0.0	0.727	0.0	0.0	0.0
Writ. Exp. 1	0.0	0.0	0.739	0.0	0.0
Writ. Exp. 2	0.0	0.0	0.787	0.0	0.0
Writ. Exp. 3	0.0	0.0	0.830	0.0	0.0
Vocab. 1	0.0	0.0	0.0	0.832	0.0
Vocab. 2	0.0	0.0	0.0	0.820	0.0
Vocab. 3	0.0	0.0	0.0	0.797	0.0
Read. Comp. 1	0.0	0.0	0.0	0.0	0.798
Read. Comp. 2	0.0	0.0	0.0	0.0	0.751
Read. Comp. 3	0.0	0.0	0.0	0.0	0.740

## Chinese Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.879	0.0	0.0	0.0	0.0
List. Comp. 2	0.861	0.0	0.0	0.0	0.0
List. Comp. 3	0.817	0.0	0.0	0.0	0.0
Struc. 1	0.0	0.697	0.0	0.0	0.0
Struc. 2	0.0	0.692	0.0	0.0	0.0
Writ. Exp. 1	0.0	0.0	0.660	0.0	0.0
Writ. Exp. 2	0.0	0.0	0.684	0.0	0.0
Writ. Exp. 3	0.0	0.0	0.758	0.0	0.0
Vocab. 1	0.0	0.0	0.0	0.791	0.0
Vocab. 2	0.0	0.0	0.0	0.770	0.0
Vocab. 3	0.0	0.0	0.0	0.768	0.0
Read. Comp. 1	0.0	0.0	0.0	0.0	0.779
Read. Comp. 2	0.0	0.0	0.0	0.0	0.754
Read. Comp. 3	0.0	0.0	0.0	0.0	0.767

Farsi Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.918	0.0	0.0	0.0	0.0
List. Comp. 2	0.877	0.0	0.0	0.0	0.0
List. Comp. 3	0.854	0.0	0.0	0.0	0.0
Struc. 1	0.0	0.771	0.0	0.0	0.0
Struc. 2	0.0	0.681	0.0	0.0	0.0
Writ. Exp. 1	0.0	0.0	0.772	0.0	0.0
Writ. Exp. 2	0.0	0.0	0.811	0.0	0.0
Writ. Exp. 3	0.0	0.0	0.813	0.0	0.0
Vocab. 1	0.0	0.0	0.0	0.829	0.0
Vocab. 2	0.0	0.0	0.0	0.827	0.0
Vocab. 3	0.0	0.0	0.0	0.790	0.0
Read. Comp. 1	0.0	0.0	0.0	0.0	0.826
Read. Comp. 2	0.0	0.0	0.0	0.0	0.757
Read. Comp. 3	0.0	0.0	0.0	0.0	0.802

French Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.906	0.0	0.0	0.0	0.0
List. Comp. 2	0.924	0.0	0.0	0.0	0.0
List. Comp. 3	0.902	0.0	0.0	0.0	0.0
Struc. 1	0.0	0.807	0.0	0.0	0.0
Struc. 2	0.0	0.758	0.0	0.0	0.0
Writ. Exp. 1	0.0	0.0	0.753	0.0	0.0
Writ. Exp. 2	0.0	0.0	0.691	0.0	0.0
Writ. Exp. 3	0.0	0.0	0.791	0.0	0.0
Vocab. 1	0.0	0.0	0.0	0.814	0.0
Vocab. 2	0.0	0.0	0.0	0.796	0.0
Vocab. 3	0.0	0.0	0.0	0.746	0.0
Read. Comp. 1	0.0	0.0	0.0	0.0	0.810
Read. Comp. 2	0.0	0.0	0.0	0.0	0.840
Read. Comp. 3	0.0	0.0	0.0	0.0	0.875

Indonesian Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.864	0.0	0.0	0.0	0.0
List. Comp. 2	0.865	0.0	0.0	0.0	0.0
List. Comp. 3	0.790	0.0	0.0	0.0	0.0
Struc. 1	0.0	0.645	0.0	0.0	0.0
Struc. 2	0.0	0.758	0.0	0.0	0.0
Writ. Exp. 1	0.0	0.0	0.634	0.0	0.0
Writ. Exp. 2	0.0	0.0	0.721	0.0	0.0
Writ. Exp. 3	0.0	0.0	0.684	0.0	0.0
Vocab. 1	0.0	0.0	0.0	0.786	0.0
Vocab. 2	0.0	0.0	0.0	0.759	0.0
Vocab. 3	0.0	0.0	0.0	0.725	0.0
Read. Comp. 1	0.0	0.0	0.0	0.0	0.737
Read. Comp. 2	0.0	0.0	0.0	0.0	0.747
Read. Comp. 3	0.0	0.0	0.0	0.0	0.722

Japanese Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.897	0.0	0.0	0.0	0.0
List. Comp. 2	0.881	0.0	0.0	0.0	0.0
List. Comp. 3	0.859	0.0	0.0	0.0	0.0
Struc. 1	0.0	0.766	0.0	0.0	0.0
Struc. 2	0.0	0.761	0.0	0.0	0.0
Writ. Exp. 1	0.0	0.0	0.695	0.0	0.0
Writ. Exp. 2	0.0	0.0	0.782	0.0	0.0
Writ. Exp. 3	0.0	0.0	0.789	0.0	0.0
Vocab. 1	0.0	0.0	0.0	0.843	0.0
Vocab. 2	0.0	0.0	0.0	0.801	0.0
Vocab. 3	0.0	0.0	0.0	0.784	0.0
Read. Comp. 1	0.0	0.0	0.0	0.0	0.815
Read. Comp. 2	0.0	0.0	0.0	0.0	0.816
Read. Comp. 3	0.0	0.0	0.0	0.0	0.792

Korean Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.885	0.0	0.0	0.0	0.0
List. Comp. 2	0.900	0.0	0.0	0.0	0.0
List. Comp. 3	0.861	0.0	0.0	0.0	0.0
Struc. 1	0.0	0.767	0.0	0.0	0.0
Struc. 2	0.0	0.737	0.0	0.0	0.0
Writ. Exp. 1	0.0	0.0	0.686	0.0	0.0
Writ. Exp. 2	0.0	0.0	0.755	0.0	0.0
Writ. Exp. 3	0.0	0.0	0.763	0.0	0.0
Vocab. 1	0.0	0.0	0.0	0.795	0.0
Vocab. 2	0.0	0.0	0.0	0.827	0.0
Vocab. 3	0.0	0.0	0.0	0.786	0.0
Read. Comp. 1	0.0	0.0	0.0	0.0	0.830
Read. Comp. 2	0.0	0.0	0.0	0.0	0.802
Read. Comp. 3	0.0	0.0	0.0	0.0	0.813

Spanish Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.899	0.0	0.0	0.0	0.0
List. Comp. 2	0.907	0.0	0.0	0.0	0.0
List. Comp. 3	0.886	0.0	0.0	0.0	0.0
Struc. 1	0.0	0.743	0.0	0.0	0.0
Struc. 2	0.0	0.723	0.0	0.0	0.0
Writ. Exp. 1	0.0	0.0	0.786	0.0	0.0
Writ. Exp. 2	0.0	0.0	0.786	0.0	0.0
Writ. Exp. 3	0.0	0.0	0.815	0.0	0.0
Vocab. 1	0.0	0.0	0.0	0.765	0.0
Vocab. 2	0.0	0.0	0.0	0.779	0.0
Vocab. 3	0.0	0.0	0.0	0.692	0.0
Read. Comp. 1	0.0	0.0	0.0	0.0	0.811
Read. Comp. 2	0.0	0.0	0.0	0.0	0.788
Read. Comp. 3	0.0	0.0	0.0	0.0	0.830

Thai Group

<u>Parcel</u>	<u>Factor</u>				
	1	2	3	4	5
List. Comp. 1	0.874	0.0	0.0	0.0	0.0
List. Comp. 2	0.881	0.0	0.0	0.0	0.0
List. Comp. 3	0.835	0.0	0.0	0.0	0.0
Struc. 1	0.0	0.542	0.0	0.0	0.0
Struc. 2	0.0	0.624	0.0	0.0	0.0
Writ. Exp. 1	0.0	0.0	0.615	0.0	0.0
Writ. Exp. 2	0.0	0.0	0.709	0.0	0.0
Writ. Exp. 3	0.0	0.0	0.747	0.0	0.0
Vocab. 1	0.0	0.0	0.0	0.785	0.0
Vocab. 2	0.0	0.0	0.0	0.736	0.0
Vocab. 3	0.0	0.0	0.0	0.756	0.0
Read. Comp. 1	0.0	0.0	0.0	0.0	0.786
Read. Comp. 2	0.0	0.0	0.0	0.0	0.727
Read. Comp. 3	0.0	0.0	0.0	0.0	0.750

# TOEFL Research Reports currently available...

- Report 1.** *The Performance of Native Speakers of English on the Test of English as a Foreign Language*. John L. D. Clark. November 1977.
- Report 2.** *An Evaluation of Alternative Item Formats for Testing English as a Foreign Language*. Lewis W Pike June 1979
- Report 3.** *The Performance of Non-Native Speakers of English on TOEFL and Verbal Aptitude Tests*. Paul J. Angelis, Spencer S. Swinton, and William R. Cowell. October 1979
- Report 4.** *An Exploration of Speaking Proficiency Measures in the TOEFL Context*. John L. D. Clark and Spencer S. Swinton. October 1979
- Report 5.** *The Relationship between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language*. Donald E. Powers. December 1980.
- Report 6.** *Factor Analysis of the Test of English as a Foreign Language for Several Language Groups*. Spencer S. Swinton and Donald E. Powers. December 1980.
- Report 7.** *The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings*. John L.D. Clark and Spencer S. Swinton. December 1980.
- Report 8.** *Effects of Item Disclosure on TOEFL Performance*. Gordon A. Hale, Paul J. Angelis, and Lawrence A. Thibodeau. December 1980.
- Report 9.** *Item Performance Across Native Language Groups on the Test of English as a Foreign Language*. Donald L. Alderman and Paul W. Holland. August 1981.
- Report 10.** *Language Proficiency as a Moderator Variable in Testing Academic Aptitude*. Donald L. Alderman. November 1981.
- Report 11.** *A Comparative Analysis of TOEFL Examinee Characteristics, 1977-1979*. Kenneth M. Wilson. July 1982.
- Report 12.** *GMAT and GRE Aptitude Test Performance in Relation to Primary Language and Scores on TOEFL*. Kenneth M. Wilson. July 1982
- Report 13.** *The Test of Spoken English as a Measure of Communicative Ability in the Health Professions: Validation and Standard Setting*. Donald E. Powers and Charles W. Stansfield. January 1983.
- Report 14.** *A Manual for Assessing Language Growth in Instructional Settings*. Spencer S. Swinton. February 1983.
- Report 15.** *Survey of Academic Writing Tasks Required of Graduate and Undergraduate Foreign Students*. Brent Bridgeman and Sybil Carlson. September 1983.
- Report 16.** *Summaries of Studies Involving the Test of English as a Foreign Language, 1963-1982*. Gordon A. Hale, Charles W. Stansfield, and Richard P. Duran. February 1984.
- Report 17.** *TOEFL from a Communicative Viewpoint on Language Proficiency: A Working Paper*. Richard P. Duran, Michael Canale, Joyce Penfield, Charles W. Stansfield, and Judith E. Liskin-Gasparro. February 1985.
- Report 18.** *A Preliminary Study of Raters for the Test of Spoken English*. Isaac I. Bejar. February 1985.
- Report 19.** *Relationship of Admission Test Scores to Writing Performance of Native and Nonnative Speakers of English*. Sybil B. Carlson, Brent Bridgeman, Roberta Camp, and Janet Waanders. August 1985.
- Report 20.** *A Survey of Academic Demands Related to Listening Skills*. Donald E. Powers. December 1985.
- Report 21.** *Toward Communicative Competence Testing: Proceedings of the Second TOEFL Invitational Conference*. Charles W. Stansfield. May 1986.
- Report 22.** *Patterns of Test Taking and Score Change for Examinees Who Repeat the Test of English as a Foreign Language*. Kenneth M. Wilson. January 1987.
- Report 23.** *Development of Cloze-Elide Tests of English as a Second Language*. Winton Manning. April 1987.
- Report 24.** *A Study of the Effects of Item Option Rearrangement on the Listening Comprehension Section of the Test of English as a Foreign Language*. Marna Golub-Smith. August 1987.
- Report 25.** *The Interaction of Student Major-Field Group and Text Content in TOEFL Reading Comprehension*. Gordon A. Hale. January 1988.



BEST COPY AVAILABLE

5790621 • HP28P3 • 275622 • Printed in U.S.A.